Capturing Complexity: Transferable flood impact models with Machine Learning

Wagenaar, D.J.

2020

VRIJE UNIVERSITEIT

Capturing Complexity: Transferable flood impact models with Machine Learning

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof. dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op woensdag 2 september 2020 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Dennis Johannes Wagenaar

geboren te Hengelo (O)

promotor: prof.dr. J.C.J.H. Aerts

copromotoren: dr. L.M. Bouwer
dr. H. de Moel

Independent members doctorate committee

prof. dr. B. Merz	GFZ Helmholtz Centre Potsdam
prof. dr. ir. M. Kok	Delft University of Technology
prof. dr. A.H. Weerts	Wageningen University
dr. M.J.C. van den Homberg	510, the Netherlands Red Cross
dr. D. Coumou	VU University Amsterdam
dr. G. Guimarães Nobre	VU University Amsterdam

Cover photo: Shutterstock.com
Printed by RODI media zh

ISBN: 978-90-8140-670-3

This thesis is based on five peer-reviewed journal articles:

Title: Uncertainty in flood damage estimates and its potential effect on investment decisions
First author: Dennis Wagenaar
Co-authors: Karin de Bruijn, Laurens Bouwer, Hans de Moel
Journal: Natural Hazards and Earth System Sciences

Title: Multi-variable flood damage modelling with limited data using supervised learning approaches
First author: Dennis Wagenaar
Co-authors: Jurjen de Jong, Laurens Bouwer
Journal: Natural Hazards and Earth System Sciences

Title: Regional and Temporal Transferability of Multivariable Flood Damage Models
First author: Dennis Wagenaar
Co-authors: Stefan Lüdtke, Kai Schröter, Laurens Bouwer, Heidi Kreibich
Journal: Water Resources Research

Title: Improved transferability of data-driven damage models through sample selection bias correction
First author: Dennis Wagenaar
Co-authors: Tiaravanni Hermawan, Marc van den Homberg, Jeroen Aerts, Heidi Kreibich, Hans de Moel, Laurens Bouwer
Journal: Risk Analysis

Title: Invited Perspectives: How machine learning will change flood risk and impact assessments
First author: Dennis Wagenaar
Co-authors: Alex Curran, Mariano Balbi, Alok Bhardwaj, Robert Soden, Emir Hartato, Gizem Mestav Sarica, Laddaporn Ruangpan, Giuseppe Molinario, David Lallemant
Journal: Natural Hazards and Earth System Sciences

TABLE OF CONTENTS

SUMMARY

Globally, floods cause, yearly, on average, an economic damage of US$ 66 billion, around 14,000 fatalities, and leave 200 million people injured, displaced or in need of emergency assistance (UNISDRR, 2017). Population growth and climate change are expected to increase these impacts in the future. To develop strategies to deal with these natural hazards, it is important to predict and quantify accurately the impacts from flood events. In flood management, such quantified impacts are needed for evaluating risk-reduction measures, for providing impact forecasts just before or after the event, and for setting (re-)insurance premiums.

Flood impacts are typically modeled as a function of the inundation depth using so-called damage curves, often designed by experts. For many areas around the world, there are no damage curves available, or they are based on outdated data. It is common practice in such cases that damage curves are transferred from other areas (e.g. using a Dutch damage curve to model flood damages in Italy). In recent years, however, model comparisons have revealed considerable differences between flood-damage curves. These large differences indicate that damage processes are too complex to be captured by a single variable, such as water depth. Therefore, when a damage curve is transferred to another location, it lacks important contextual information, such as early warning, flood experience, and building style, all of which may influence the damage. Thus, transferring damage models causes bias errors, which are systematic over- or underestimations of the damage.

In this thesis, a new approach for developing transferable damage models is explored using data-driven methods. This approach is applied and tested to model damage in transfer settings. Multi-variable flood-damage models are developed using machine-learning (ML) techniques and data from past flood-damage. These damage data are upgraded with additional variables to improve the predictive capacity of the resulting models. These data-driven models are then transferred outside their original geographical context to a different location. Domain-adaptation techniques to correct for the sample-selection bias are applied to improve the ML techniques. Three different datasets are applied: damage data from the floods in the Meuse Basin in 1993 (Netherlands); HOWAS data on German flood damages from several events between 2002 and 2013; and impact data from 12 typhoons in the Philippines between 2012 and 2016. Models are transferred between Germany and the Netherlands and among different places in the Philippines.

This thesis demonstrates that both ML and domain-adaptation techniques work well for reducing bias errors in flood-damage models, with reductions in mean bias error of about 50%. The approach is, therefore, sometimes suitable for transferring models to areas in which no effective local model is available. The ML techniques work less well for reducing absolute errors concerning the randomness at specific predictions. Therefore, for precise predictions on property level, the techniques provide an improvement but no leap forward. However, when more and better data about buildings and historical damage are available (e.g. through remote sensing techniques), an even larger improvement can be expected with the techniques investigated in this thesis.

SAMENVATTING

Wereldwijd veroorzaken overstromingen jaarlijks gemiddeld een economische schade van 66 miljard dollar, ongeveer 14.000 dodelijke slachtoffers, en worden 200 miljoen mensen op een andere manier getroffen (e.g. gewond, ontheemd of noodhulp nodig) (UNISDRR, 2017). Bevolkingsgroei en klimaatverandering zullen naar verwachting deze effecten in de toekomst vergroten. Om strategieën te ontwikkelen om met overstromingen om te gaan, is het belangrijk om de effecten van overstromingen nauwkeurig te voorspellen en te kwantificeren. Bij overstromingsbeheer zijn dergelijke gekwantificeerde effecten nodig voor het evalueren van risicobeperkende maatregelen, voor het maken van impactvoorspellingen vlak voor of na de ramp en voor het vaststellen van (her) verzekeringspremies.

Overstromingen worden typisch gemodelleerd als functie van de waterdiepte met behulp van zogenaamde schadefuncties, vaak ontworpen door experts. Wereldwijd zijn er voor veel gebieden geen schadefuncties beschikbaar of zijn deze gebaseerd op verouderde gegevens. In dergelijke gevallen is het gebruikelijk dat schadefuncties uit andere gebieden worden toegepast (bijvoorbeeld met een Nederlandse schadefunctie overstromingsschade in Italië modelleren). In de afgelopen jaren hebben modelvergelijkingen echter aanzienlijke verschillen tussen schadefuncties aan het licht gebracht. Deze grote verschillen geven aan dat schadeprocessen te complex zijn om door één enkele variabele, zoals waterdiepte, te worden vastgelegd. Wanneer een schadefunctie in een andere locatie wordt toegepast, mist het daarom belangrijke contextuele informatie, zoals de kwaliteit van de waarschuwing, ervaringen met overstromingen van de lokale bevolking en de lokale bouwstijl, die allemaal de schade kunnen beïnvloeden. Het elders inzetten van schademodellen veroorzaakt dus zogenaamde biasfouten, dat zijn systematisch over- of onderschattingen van de schade.

In dit proefschrift wordt een nieuwe aanpak voor het ontwikkelen van schademodellen verkend die beter toepasbaar zijn buiten hun initiële context, hiervoor worden data-gedreven methoden gebruikt. Deze aanpak wordt toegepast en getest om schade te modelleren in een situatie waarbij een model buiten zijn initiële context wordt toegepast. Multi-variabele overstromingsschade-modellen worden ontwikkeld met behulp van machine learning (ML) technieken en gegevens van eerdere overstromingsschade. Deze schadegegevens worden geüpgraded met aanvullende variabelen om de voorspellende capaciteit van de resulterende modellen te verbeteren. Deze datagestuurde modellen worden vervolgens buiten

hun oorspronkelijke geografische context in een andere locatie toegepast. Zogenaamde “Domain adaptation” technieken worden gebruikt om de ML-technieken te corrigeren voor de fout die wordt veroorzaakt doordat de data waarmee het model is gemaakt niet geheel representatief zijn voor de gebieden waarin het wordt toegepast. Er worden drie verschillende datasets toegepast: schadegegevens van de overstromingen in van de Maas in 1993 (Nederland), HOWAS-gegevens over Duitse overstromingsschade door verschillende overstromingen tussen 2002 en 2013 en impactgegevens van 12 tyfoons in de Filipijnen tussen 2012 en 2016. Modellen worden overgedragen tussen Duitsland en Nederland en tussen verschillende plaatsen in de Filipijnen.

Dit proefschrift toont aan dat zowel ML- als domain adaptation technieken goed werken voor het verminderen van biasfouten in overstromingsschade-modellen, met een reductie van de gemiddelde biasfout van ongeveer 50%. De aanpak is daarom soms geschikt om modellen over te dragen naar gebieden waar geen effectief lokaal model beschikbaar is. De ML-technieken werken minder goed om absolute fouten met betrekking tot specifieke gebouwen te verminderen. Daarom, voor nauwkeurige voorspellingen op gebouwniveau, bieden de technieken slechts een kleine verbetering. Als er echter meer en betere gegevens over gebouwen en historische schade beschikbaar komen (bijvoorbeeld door middel van remote sensing technieken), kan een nog grotere verbetering worden verwacht met de technieken die in dit proefschrift zijn onderzocht.

1. INTRODUCTION

In November 2013, Typhoon Haiyan, one of the strongest typhoons ever recorded, was heading for Tacloban in the Philippines. Warnings were issued to the several hundred thousand residents of the coastal city. They had experienced typhoons before, but few could imagine what was about to happen. The direction and the force of the typhoon caused a four-meter-high storm surge and the associated flooding, which came as a surprise to most people. The record-breaking wind speeds combined with the storm surge caused high waves and strong currents in the flood water. People vanished as their houses collapsed in the swirling flood waters or drowned trapped inside because of the rising water. When the storm calmed down and the flood water receded, the streets were left littered with dead bodies. Over 6000 people died and all the buildings were either destroyed or severely damaged.

Globally, natural hazards such as Typhoon Haiyan cause, yearly, on average, economic damage of US$ 291 billion, kill 130,000 and leave 440 million people injured, displaced, or in need of emergency assistance (UNISDRR, 2017). Floods alone, yearly, cause US$66 billion in damages, 14,000 fatalities and leave 200 million people injured, displaced, or in need of emergency assistance (UNISDRR, 2017). Without improvements in disaster risk reduction, these numbers are likely to increase because many natural hazards will become more frequent due to climate change (Kundzewicz et al., 2014) and it is expected that the number of people living in at-risk areas will increase (Bouwer, 2012). Low-income communities are especially affected by natural hazards, with a poor person being six times more likely to be affected than a rich person (UNISDRR, 2017). For some countries, the average annual economic damage is a significant share of their Gross Domestic Product (GDP), which is a significant hindrance in the economic development of poor countries (UNISDRR, 2017).

It is possible to prevent meteorological events such as Typhoon Haiyan from turning into major disasters. For example, flood protection structures could have prevented the storm surge from causing severe flooding, better warning systems may have helped to evacuate more people to safer locations, and appropriate insurance products may have been able to help the households and businesses to recover more easily. To enable such measures, predictive information and knowledge on disaster impacts (e.g. economic damage, number of buildings destroyed, number of casualties, number of people affected) is required. This knowledge needs to come from predictive models that capture complex combinations of physical and socio-economic processes that cause such impacts. The purpose of this PhD thesis is to improve such predictive natural-hazard impact models by applying data and ML to help reduce the negative impacts of disasters despite climate change and population growth augmenting these impacts.

1.1 Natural-hazard impact models

1.1.1 Definitions and types of impacts and damage

Impacts from natural hazards include both negative, neutral and positive consequences. The terms “loss” and “damage” are typically applied for the negative impacts of a disaster. Loss and damage are often applied interchangeably, but sometimes also to differentiate between physical damage and economic loss (e.g. Koks, 2016). In that case, economic loss is the non-physical damage. In this thesis, the term “loss” is not applied because this thesis focuses only on physical damage and to avoid any possible confusion.

Natural-hazard damage is often categorized as tangible or intangible and direct or indirect (Table 1.1). Tangible damage is all the damage that can be expressed in monetary terms; whereas, intangible damage cannot be easily expressed in monetary terms (e.g. casualties). Direct flood damage is damage within the flooded area, and indirect flood damage is outside the flooded area (e.g. through network effects) (Jonkman et al., 2008).[1] Direct tangible damage can be split into physical damage and direct interruption damage. Direct interruption damage is the interruption of physical assets (e.g. building functions) directly affected by the natural hazard (i.e. damaged buildings cannot be used); whereas, indirect interruption damage affects buildings or activities outside the area that is physically affected by the natural hazard (i.e. supplier damaged so company cannot operate) (Meyer et al., 2013). Most flood-damage models only focus on direct tangible physical damage. The primary reason for this focus is the relative difficulty of modeling the many other types of damage. However, these other types of flood damage can be a considerable share of the total flood damage, depending on the case (Gauderis & Kind, 2011; Vilier, 2013). In the case of the Dutch national damage models, direct physical damage is assumed to be the largest portion of the total flood damage, and this is even the case when compared with all other damage types, including monetized intangible flood damage (Gauderis & Kind, 2011). However, this

[1] Some other definitions are also common in which business interruption inside the flooded area is considered indirect damage (e.g. Merz et al., 2010).

is not always the case; Koks et al. (2016) argue that, in some cases, with small flood probabilities, non-physical damage can be greater than physical damage.

Table 1.1: Overview of different flood-damage types (Jonkman et al., 2008)

	Tangible	***Intangible***
Direct	• Residences • Capital assets and inventory • Business/housing services interruption (inside flooded area) • Vehicles • Agriculture land and cattle • Roads, utility and communication infrastructure • Evacuation and rescue operations • Reconstruction of flood defenses • Clean-up costs	• Fatalities • Injuries • Inconvenience and moral damage • Utilities and communication • Historical and cultural losses • Environmental losses
Indirect	• Damage for companies outside the flooded area • Adjustments in production and consumption patterns outside the flooded area • Temporary housing of evacuees	• Societal disruption • Psychological trauma • Undermined trust in public authorities

Depending on the purpose of the impact estimates, the same impacts can be accounted for in different ways. For example, damages to residences are sometimes expressed in monetary terms, but also often in number of buildings damaged or destroyed. It is sometimes also useful to express disaster impacts in terms of needs. For example, the number of needed tents, building materials, medicine or food supplies. These needs are implicitly already part of some of the other damages included in Table 1, but, for some applications (especially impact forecasting), it is beneficial to express them more narrowly as “needs” rather than “damage.”

1.1.2 Modeling damages

Direct physical flood damage is traditionally estimated based on a relationship between a hazard indicator (usually water depth) and damage for different types of buildings (Merz et al., 2010; Scawthorn et al., 2006; Penning-Rowsell, 2005; Dias et al., 2018). These relationships are called damage functions and were first proposed by White (1945). They can either be applied on a large number of objects at the same time or at individual objects (called the unit-loss method). There are currently two techniques applied in developing flood-damage functions, which are often combined:

1. Synthetic approach
2. Data-driven approach

The synthetic approach is based on a set of “what if” questions, such as “what damage would you expect if the flood is 2m above the floor?” (Merz et al., 2010). This approach is sometimes applied in great detail, with questions being asked about each specific element in a building (e.g. door, window, etc.) (e.g. Penning-Rowsell et al., 2005). The United States Army Corps of Engineers (USACE) has a long history of developing highly detailed damage functions using this method for a large range of different buildings and circumstances (USACE, 1997; USACE, 2006; USACE, 2015). Answers to “what if” questions are sometimes supported with expert calculations (e.g. Vrouwenvelder & Waarts, 1994; Roos, 2003). In these calculations, assumptions are made about flood conditions and building characteristics. A mechanical calculation is then typically carried out to check whether the building will collapse with the given assumptions. Such calculations have been carried out for hydrostatic action on buildings (Vrouwenvelder & Waarts, 1994) and hydrodynamic actions on buildings, such as flow velocity (Roos, 2003) and waves (Roos, 2003), and even types of debris action, such as tree trunks floating into a building (Roos, 2003).

A different method of damage modeling is through data-driven approaches (Merz et al., 2010). These data consist of variables that can be used to predict the damage (e.g. water depth or flow velocity) and a corresponding observed damage (e.g. Islam, 1997; Pistrika&Jonkman, 2010). A regression analysis is then carried out to find a relationship between the predictor variables, such as water depth or flow velocity, and damage. However, these traditional regression techniques seem unable to capture the full complexity of the damage processes. These techniques also require a synthetic component because the damage modeler needs to choose the variables for the regression, which implies a preconceived damage mechanism, a proper regression function, potentially transforming some of the variables, and introduces simplifications.

More recently, traditional regression analysis techniques have been replaced by more advanced ML techniques (Merz et al., 2013). This is a a more fully data-driven approach for flood-damage modeling because these newer algorithms can in theory better capture the potentially non-linear relationships across many variables behind the damage processes. Furthermore, no choice of regression function needs to be made and no variables need to be transformed. Therefore, the damage modeler needs less insight into the precise physical processes. This makes a fully data-driven approach different from a traditional regression approach, which is based on assumptions about the major damaging processes. This could make the models cheaper and easier to transfer but could also be a disadvantage because the modeller will understand the results less well. Making use of and improving such data-driven approaches is the purpose of this PhD thesis.

In practice, some elements of a data-driven approach typically end up in a synthetic approach as well. Usually, this usage is indirect, for example, because the expert estimates are based on empirical observations by another expert in the past. Sometimes, this usage is more formal, such as an expert referencing data for their estimates. An illustrative example of this usage is the Dutch standard method for modeling flood damages (Duiser, 1982; Kok et al., 2005; De Bruijn et al., 2014). This method is based on empirical observations from several floods, including the Wieringermeer flood of 1945, the large coastal floods in Zeeland in 1953, and floods in Tuindorp-Oostzaan in 1960. Furthermore, experts were interviewed, multiple foreign damage models were considered, and expert calculations were carried out (Duiser, 1982) to form the Dutch standard method. The experts combined all this information into depth-damage functions.

1.2 Application of impact models

In this section, different applications of impact models are discussed, as well as the need for more accurate and reliable damage models. Merz et al. (2010) defines six applications for flood damage modeling: assessment of vulnerability, flood risk mapping, optimal decisions on flood mitigation measures, comparative risk analysis, financial appraisals for the (re-)insurance sector, and financial appraisals during and immediately after floods. The techniques developed and investigated in this thesis are applicable to all six of these applications.

For this section assessment of vulnerability, flood risk mapping, optimal decisions on flood mitigation measures, and comparative risk analysis are grouped together as "Mitigation: assessment of permanent risk reduction measures" because they share the goal of reducing flood risk in the mitigation stage of the disaster-management cycle (Khan et al., 2008; National Research Council, 2006). Then, what are labeled "financial appraisals" during and immediately after floods in Merz et al. (2010) are discussed as "Preparedness: impact-based forecasting for temporary risk reduction measures," and the section ends by considering the distinct applications for the (re-)insurance sector where models are applied for risk transfer rather than for risk reduction (Van den Homberg & McQuistan, 2019). An overview of the different applications is shown in table 1.2.

Table 1.2 Overview of the different applications of flood damage models, ordered by the position in the DRM Cycle (Khan et al., 2008; National Research Council, 2006) and the risk objective (Van den Homberg & McQuistan, 2019)

	Mitigation and prevention	*Preparedness*	*Response*	*Recovery*
Risk Reduction	DRR-risk reduction: assessment of vulnerability, flood risk mapping; optimal decisions of flood mitigation measures, comparative risk analysis.	DRR-impact-based forecasting; early warning early action		
Risk Transfer		(Re-)Insurance	(Re-)Insurance	(Re-)Insurance

1.2.1 Mitigation: assessment of permanent risk reduction measures

All risk-reduction measures come at a cost, and to identify those measures that are worth the cost, predictive impact models are applied. A common approach is to calculate the impact for many different natural-hazard scenarios (e.g. different water levels), with and without the measure. The reduction of impacts in all the different scenarios and the probability of those scenarios are then applied to determine the expected benefits of the measures (e.g. Aerts et al., 2014). This benefit can then be compared with the cost of the measures in a cost-benefit analysis.

This approach can be applied in several ways. It can be used for highly detailed studies that evaluate specific measures, for example, evaluating whether widening a drainage canal is worth the cost (e.g. Wagenaar et al., 2019). A more advanced method is to address the dimension of measures as well, for example, the protection level of a dike (e.g. Kind, 2011; van der Most et al., 2014). This method is used by repeating the cost-benefit analysis for different designs and assessing the net benefit of the different options. It can be applied to packages of measures also, often referred to as "adaptation strategies" or "adaptation pathways" (Ruig et al., 2019).[1] Measures can take many different forms, and cost-benefit analyses can be made for either risk-reduction projects with specific physical measures (e.g. Wagenaar et al., 2019) or for a set of policy regulations. Such regulations can take the form of building codes (e.g. De Ruig et al., 2019)[2] or safety standards for dikes (e.g. Kind, 2011). For example, safety standards are applied to the minimum strength of buildings to protect them from wind or earthquakes, or the required building elevation to protect them from floods (Ruig et al., 2019).[2]

Sometimes, it is useful to obtain a general order of magnitude of the expected disaster impacts to compare the risk in different regions and for different natural-hazard types (e.g. Grunthal et al., 2006). Impact maps are also used in (spatial-) planning and risk maps to guide zoning regulations and urban development. Such a risk map uses the information from impact models to reveal the expected flooding (-impacts) at a location (De Bruijn et al., 2015). Risk maps can be either directly used by spatial planners to determine suitable locations for new urban areas, or they can be applied to set protection standards for levees. For example, the future standards for Dutch dikes will be such that, in the protected areas, the probability of dying from flooding will always be smaller than 10^{-5} per year (Van der Most et al., 2014).

The required precision of damage models for assessing risk-reduction measures depends on the purpose for which the models are being applied. Especially when measures reduce the damage component, rather than only the probability component of the risk, an accurate and reliable model is crucial. This is the case when, for instance, water depths are reduced (e.g. Wagenaar et al., 2019), or when proposed flood-proofing measures are tailored to the building scale. In this second case it is required to include information on building characteristics, various hydrodynamic parameters, and socio-economic conditions to judge the effect of the measure (Richert et al., 2019).

1.2.2 Preparedness: impact-based forecasting for temporary risk reduction measures

Warning systems are a common approach to reduce disaster risk (Coughlan de Perez et al., 2015). Such systems have traditionally been mostly concerned with weather forecasts or water level predictions (Verkade & Werner, 2011). Recently, these warning systems have been evolving from predicting “what the weather will be” to “what the weather will do” (Guimarães Nobre, 2020). The main reason for this evolution is that a warning about what the weather will do is more easily understood and acted upon by the public, especially when the warning concerns a disaster that does not frequently occur. A second reason is that this system will lead to better-informed decisions, and such impact-based warning information could be used as a first estimate to complement relief and recovery needs assessments.

Based on a forecast, early actions that can be carried out before a disaster strikes are often triggered automatically by specific forecasted levels. These forecasted levels can be water related (e.g. a storm surge barrier closes because of predicted water level), but they can also be impact related (e.g. trigger based on a certain number of predicted destroyed buildings). Such early actions based on forecasts can be more general, such as to initiate an evacuation or to provide emergency food supplies (Coughlan de Perez et al., 2014). Some of the early actions are specifically targeted, for example, reinforcing vulnerable buildings (Coughlan de Perez et al., 2014), cash transfers to specific households (Reuters, 2019), or early harvesting of crops. These specifically targeted early actions may especially benefit from impact forecasts. For an effective execution of these actions, reliable damage estimates are important. However, the uncertainty in an early forecast can be large, and the cost of acting in vain can be large as well (Bischiniotis et al., 2019). Reducing the uncertainty in the impact forecast is, therefore, crucial (Bischiniotis et al., 2020), and a reliable damage model may be one of the areas to reduce such uncertainties in an impact forecast. This is especially the case for more targeted actions in which decisions depend more on the damage forecast and in which the costs of making wrong decisions are larger. Given the uncertainty in current damage models (e.g. Jongman et al., 2012), more accurate damage models are required for these targeted early actions. Bischiniotis et al. (2020) showed that in the comparison between a forecasting system and permanent risk reduction measures the forecast quality is crucial.

1.2.3 Preparedness and response: insurance

Risk-transfer mechanisms, such as insurance, are a way to cover financial damages. An insurer (an entity that provides insurance) agrees with a policyholder to cover damages in case of a disaster in return for an annual premium. Insurance schemes can take many different forms. The policyholder can be an individual, but it can also be a company or even a country. An insurance policy can be managed by private companies, public authorities, or a mix of those two entities. Especially for disaster impacts, flood damage may affect a large pool of insurance policyholders at the same time (e.g. all the residences of a flooded city). In the insurance realm, this situation is called “correlated risk”, which poses a large risk for the financial viability of an insurer. Therefore, primary insurers cover extreme risk with re-insurance companies to protect them from very large losses.

Payments by insurers typically depend on the damages that occurred (indemnity insurance), and after the disaster, damage experts determine how much is paid to everyone. To design an economically viable insurance scheme, insurance companies require a reliable estimate of the premium for each potential policyholder. An underestimation of the impacts and related risk by the insurer could lead to less income from premiums, and, thus, a high financial cost after an event. An overestimation of the impacts could lead to high premiums, which leads to lower market penetration and lower premium earnings. Uncertainty in impact models is, therefore, negative for the insurance sector and increases the cost of an insurance policy (Kunreuther et al., 1993).

The disadvantage of the above traditional expert-based damage assessment for insurance is that it takes experts a lot of time and resources to assess who incurred how much damage after an event. An alternative, therefore, is so-called parametric insurance (Martinez-Diaz et al., 2019), in which the damage payment does not depend on the actual damage but on a predefined trigger (e.g. a storm >200mm/day) (e.g. from an independent meteorological institute) (Yuzva et al., 2018). This type of insurance can pay out much quicker and has fewer overhead costs as no detailed post disaster damage surveys are required. Forecasted damages using flood-damage modeling could play a role in this type of insurance scheme by using the forecasted impact as a trigger for the payment. Such an insurance product would need a reliable impact forecast to make sure premiums and payout reflect the objective risk.

More accurate damage estimates are, in some cases, required for the assessment of risk-reduction measures, impact forecasting and insurance. This additional accuracy can be acquired with either better data-driven approaches or better synthetic approaches. In this thesis, the focus is on the former.

1.3 Current challenges and approach

Currently, there are two challenges that cause damage models not to be accurate: (a) the damage models do not accurately capture all damage processes; and (b) damage models often need to be transferred from other areas or events because of a lack of data to develop a reliable model considering the local context.

(a) There are different processes that cause direct physical flood damage. Kelman and Spencer (2004) identified eight groups: hydrostatic, hydrodynamic, buoyancy, erosion, debris, chemical, biological, and nuclear actions. These damage processes interact with each other, and their influence also depends on asset (building, infrastructure, etc.) characteristics and their users. A traditional flood-damage model (using a depth-damage function) summarizes all this complexity in one function that depends only on the water depth (Merz et al., 2010; Molinari et al., 2019). The damage functions that result from this approach differ greatly from each other (Jongman et al., 2012). Even high-quality synthetic damage models have this problem. For example, the USACE has standardized methods to develop synthetic damage functions and records many of their assumptions (e.g. USACE, 2006). The USACE carried out two such studies in nearby areas within the state of Louisiana and came to significantly different results without being able to explain why the second study differed from the first (USACE, 2006). These studies were carried out in areas that flood regularly and by experts who had visited flood-damaged buildings in the region (USACE, 2006).

(b) Many flood-prone areas around the world do not experience flood events frequently, and no damages may have been recorded for a long time. This is especially the case for areas that are well protected or for regions that do not keep adequate records of past flood damages. For example, the last records of large-scale dike breach flooding in the west of the Netherlands date to 1953. Since then, protection levels have been the highest in the world; hence, similar large damaging floods have not occurred. In developing countries, it may be difficult to develop damage models either because data are not recorded or because of the costs of developing effective synthetic models. Therefore, in practice, damage models often need to be transferred. This transfer can be a spatial transfer (i.e. damage model originates from another area) or a temporal transfer (i.e. the model is based on an event from the past that may not be fully representative for the current situation). The problem with spatial transfers is that flood types and building types differ among areas. Moreover, people/cultures differ among areas, which may lead to different human responses across areas. Temporal transfers can cause problems because building styles change over time and the next flood may be of a different type than the last. Also circumstances can be different than in the past, for example because of consecutive disasters (De Ruiter et al., 2020). Furthermore, flood experience may cause problems for temporal transfers, for example, the Meuse floods of 1993 and 1995 in the Netherlands. These floods were very similar, they took place in the same area, and even the water depths were very similar. However, the second time, the damage was much lower because of a better flood warning and a population more experienced with flooding (Wind et al., 1999). Similar observations were made in Germany (Bubeck et al., 2012) and France (Poussin et al., 2014).

The hypothesis is that both the above problems can be solved by additional model complexity and by including more explanatory variables/complexity. This approach has already been the goal of other studies, such as Merz et al. (2013), and this thesis expands on that work. However, what is new in this thesis is the idea that additional explanatory variables (complexity) can also solve challenge (b) and help to transfer damage models more effectively. Additional model complexity can be added through a data-driven or synthetical approach. This thesis specifically explores a data-driven approach to add model complexity.

1.4 Goal and research questions

The above described hypothesis led to formulating the main goal of this thesis:

How to improve flood-damage models and make them transferable using a data-driven approach?

The question refers to the main research challenge, to illustrate that a purely data-driven approach can be applied to make complex transferable damage models. This thesis explores methods to realize this challenge, updates existing datasets, and assesses these methods to improve the quality and transferability of flood damage models.

1.4.1 Research questions

To answer the main research question, several sub-questions have been formulated and are addressed in this thesis:

Sub-question 1: What are the main sources of uncertainty in flood damage models?

To assess whether data-driven approaches improve flood damage modeling, we need to understand the uncertainty present in current flood damage models. Knowing these sources of uncertainty, one can try and reduce them using data-driven approaches.

Sub-question 2: What methods are available to develop complex multi-variable damage models?

To improve data-driven damage models, it is necessary to explore the ML methods to develop such models and to investigate methods that add additional explanatory variables to a dataset.

Sub-question 3: Do multi-variable damage models perform well when they are transferred to other locations and events?

The expectation that a multi-variable damage model that is transferred performs reasonably well compared with local models needs to be tested. These test results can provide insights into the conditions under which such a transfer is possible and can offer advice on how to improve data-driven damage models.

Sub-question 4: Are there techniques to improve the transferability of multi-variable damage models?

Damage models often need to be applied in a transfer setting. When this transfer setting is considered during the development of the data-driven damage models, it may be possible to achieve better results.

The standard methods to develop data-driven multi-variable damage models do not consider knowledge about the context in which a damage model is applied. There is a possibility to improve data-driven damage models by considering the area they will be applied in during the model training process. This question addresses this possibility.

Sub-question 5: How can the required data for data-driven impact models be acquired at scale?

A possible way to improve data-driven models is to acquire more explanatory variables. These data are, currently, often unavailable. Before the techniques explored in this thesis can be commonly applied, this data challenge needs to be resolved.

1.4.2 Introduction to the chapters

This thesis is organized into five chapters that each answer one of the sub-questions. Each chapter consists of a peer-reviewed paper. See also figure 1.1 that shows how chapter 2-5 relate to each other.

Following this introduction, the second chapter, **"Sources of uncertainty in flood damage models,"** departs from the traditional deterministic damage-curve approach. A traditional damage modeler is confronted with the problem of having many different damage functions to choose from. In this chapter, we address why these damage functions are so different from each other and what uncertainty a damage modeler should attempt to reduce for improved results. This uncertainty analysis sets a baseline that we hypothesize can be improved by data-driven approaches in the following chapters.

In the third chapter, **"Development of multi-variable flood damage models,"** the data and ML techniques to make the data-driven models are assessed. To do this, a new dataset is prepared for data-driven damage modeling and upgraded with more variables. Then, several ML methods are applied and compared with a reference damage function: multi-variable linear regression, regression trees, bagging trees, random forests, and Bayesian networks.

In the fourth chapter, **"Transferability of multi-variable flood damage models,"** the hypothesis that multi-variable damage models are transferable to other areas than where the data were collected is tested. To do this, spatial and temporal model transfers are carried out between flood events in Germany and the Netherlands.

In the fifth chapter, **"Sample selection bias correction,"** methods are introduced to improve the transferability of the multi-variable data-driven damage models developed. This task is achieved by considering the area the model will be applied in during the model training process. This approach is then proven to work for two case studies.

The sixth chapter, **"How machine learning will change flood risk and impact assessments,"** presents a literature review and viewpoint from experts on the possibilities to use ML for flood risk assessments. Special focus is placed on the methods to collect the appropriate data for data-driven damage modeling and on other types of data-driven impact models. In addition, common ML challenges are discussed.

The thesis ends with a synthesis in which the findings of the thesis are summarized and recommendations are made.

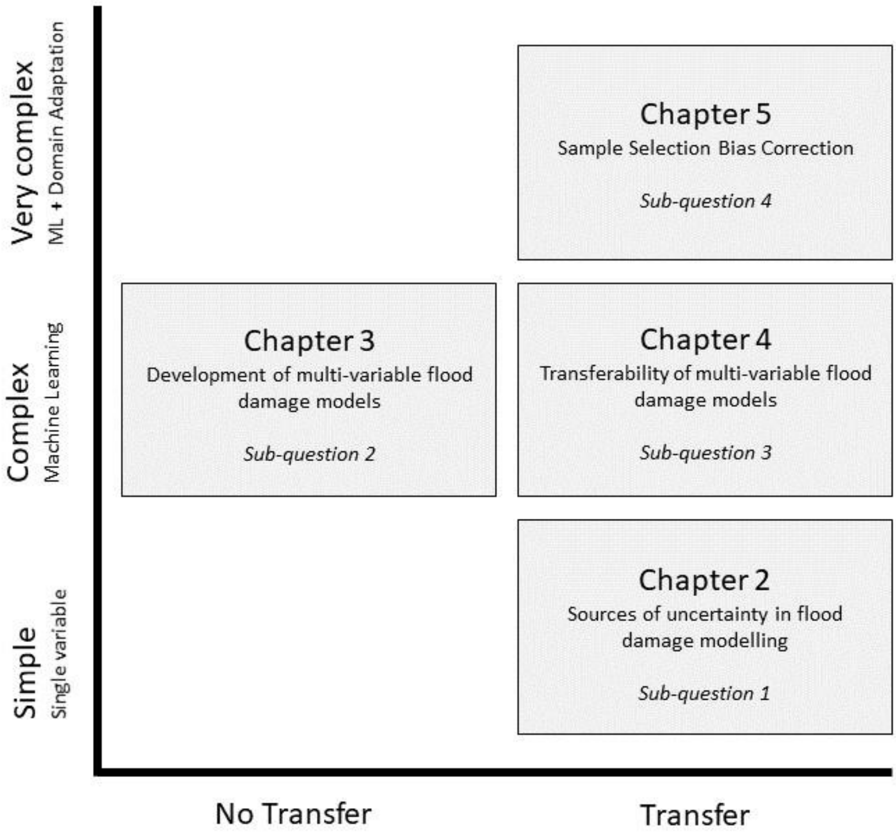


Figure 1.1. Overview of how chapter 2-5 relate to each other.

2. SOURCES OF UNCERTAINTY IN FLOOD DAMAGE MODELS

This chapter is based on the paper "Uncertainty in flood damage estimates and its potential effect on investment decisions" written together with Karin de Bruijn (Deltares), Laurens Bouwer (Climate Service Center Germany) and Hans de Moel (VU University). This is published in the journal Natural Hazards and Earth Science Systems (NHESS). The reference is: Wagenaar, D. J., de Bruijn, K. M., Bouwer, L. M., and de Moel, H.: Uncertainty in flood damage estimates and its potential effect on investment decisions, Nat. Hazards Earth Syst. Sci., 16, 1–14, https://doi.org/10.5194/nhess-16-1-2016, 2016.

Abstract

This chapter addresses the large differences that are found between damage estimates of different flood damage models. It explains how implicit assumptions in flood damage functions and maximum damages can have large effects on flood damage estimates. This explanation is then used to quantify the uncertainty in the damage estimates with a Monte Carlo Analysis. The Monte Carlo analysis uses a damage function library with 272 functions from 7 different flood damage models. The paper shows that the resulting uncertainties in estimated damages are in the order of magnitude of a factor 2 to 5. The uncertainty is typically larger for flood events with small water depths and for smaller flood events. The implications of the uncertainty in damage estimates for flood risk management are illustrated by a case study in which the economic optimal investment strategy for a dike segment in the Netherlands is determined. The case study shows that the uncertainty in flood damage estimates can lead to significant over- or under-investments.

2.1 Introduction

Flood damage assessment is an essential aspect of flood risk management (Merz et al., 2010). It is used for supporting policy analysis and flood insurance. In the Netherlands flood damage estimates are used, for example, to determine economic optimal protection standards for flood defenses (Van der Most et al., 2014), prioritize investments (Jongejan & Maaskant, 2013) or to compare the impact of different flood risk management strategies (Kind et al., 2014).

The most commonly used method for flood damage assessment is the unit loss method (De Bruijn, 2005). This method assesses the damage for each unit separately. This assessment is based on a maximum damage per object and a damage function. A damage function describes the relationship between a flood characteristic (most often water depth) and the fraction of the economic loss that occurs to the object that is damaged.

There are many different flood damage models all based on the unit-loss method (e.g. HIS-SSM for the Netherlands (Kok et al., 2005), Multi Coloured Manual in the UK (Penning-Rowsell et al., 2010), HAZUS in the USA (Scawthorn et al., 2006) and FLEMO in Germany (Thieken et al., 2008; Kreibich et al., 2010)). These models differ for good reasons. Each model is specifically derived for a specific country, region and/or flood type and tailored to characteristics of the flooding and objects in the considered region (Cammerer et al., 2013).

When these different models are applied to one and the same event, they will yield significantly different results (De Moel and Aerts, 2011; Jongman et al., 2012; Chatterton et al., 2014). Jongman et al. (2012) compared the damage outcomes of seven different flood damage models with the recorded flood damages from events in the UK and Germany. The difference between the smallest and largest estimate/recording was a factor 5 for the German event and a factor 10 for the event in the UK. Chatterton et al. (2014) compared two different damage assessments for a region in the UK. The damage estimates differed about a factor 5 to 6 for both residential and commercial damages.

These large differences in outcomes for events for which the different models all should be applicable, indicate that flood damage estimation is prone to large uncertainties and thus that potentially large errors can occur when flood damage models are applied.

The uncertainties and potential errors in damage estimates affect decision-making based on those damage estimates. A quantification of the uncertainty in the damage estimates can help to get an insight in the potential error that can occur in a decision based on the flood damage estimate and may improve the decision-making process. USACE (1992) and Peterman and Anderson (1999) both showed that taking into account ranges of uncertainty can lead to different decisions than using single value estimates.

Furthermore, uncertainty quantification is useful to expose key focus points for the improvement of flood damage estimation methods. To reduce uncertainties additional effort may be needed on research on the flood damages or on data collection of damaged objects during floods.

Previous uncertainty analysis did not show a common understanding of the size or cause of uncertainties, which also indicates that further research is needed. Generally, uncertainty in flood damage assessment is quantified with forward uncertainty propagation methods which use Monte Carlo simulations (Merz et al.,2004; Egorova et al.,2008, Apel et al.,2008; De Moel et al., 2012). The results of Egorova et al. (2008) indicate moderate uncertainties, which is in contrast with the large differences between flood damage models that were found by De Moel and Aerts (2011), Jongman et al. (2012) and Chatterton et al. (2014).

This chapter provides a method to get a robust estimate of the uncertainty in damage estimates based on an analysis of the cause of the large differences between the various existing damage models. The method is illustrated with clear hypothetical examples and then applied to a case to show its use in decision making on protection standards. The method makes use of a damage function library with 272 damage functions from 7 different flood damage models.

The paper focuses on direct material damage. Indirect damages including damages due to business interruption are not considered here, since their analysis requires different methods. The paper starts with a qualitative analysis of the uncertainty found in flood damage models. This qualitative analysis is the basis for the assumptions made in a Monte Carlo analysis which is used to quantify uncertainty. Next, the Monte Carlo analysis is described and discussed in detail. Finally, the Monte Carlo analysis is applied on a case-study in the Netherlands and the resulting uncertainties in damage estimates and the effects on flood risk management decisions are discussed.

2.2 Qualitative uncertainty analysis

This sub-chapter first provides a detailed description of unit loss flood damage models and then places all elements of such models into a framework for uncertainty classification in order to generate a detailed qualitative understanding of the uncertainty sources and effects and their correlation. The understanding is used in sub-chapter 2.3 to enable a quantitative uncertainty analysis.

2.2.1 The unit loss method for flood damage assessment

The unit loss method uses relationships between flood characteristics and damages to a unit. The unit loss method consist of four elements: The maximum damage s_i for each category, the flood characteristics (such as water depth *d) at all locations j*, the damage functions (f(d)) for all categories which determine the damage fraction and the number of objects affected (*n*). Damage of an area is assessed as the sum of all damage categories *i* for all grid cells *n by* the following formula (Egorova et al., 2008):

$$Damage = \sum_{i=1}^{m} s_i \sum_{j=1}^{n} f_{ij}(d_j) n_{ij}$$

Potentially relevant flood characteristics are the maximum water depth, flood duration, and flow velocity, pollution, warning time and other possible aspects of the flood. Often, only the water depth is used in flood damage modeling, occasionally supplemented by one or two other parameters. The uncertainties in the flood characteristics which are used as input for the damage estimation are not part of this paper.

Damage is usually calculated for categories such as houses, industries, and commercial companies, roads, and agriculture. These object categories differ in maximum damage and flood damage functions. The object and flood characteristics are linked by damage functions which give the fraction of the maximum damage which occurs as a function of the flood intensity. The damage fraction is then multiplied with the maximum damage to get the damage. (Some methods such as the one in the Multicoloured Manual (Penning-Rowsell et al., 2010) use absolute damage functions which relate the flood intensity directly to the damage and not to a fraction of the maximum damage).

The maximum damage can be defined in different ways. In this analysis we define the maximum damage as the expected damage corresponding with an extreme water depth. This means that the damage function will reach the value of one/unity for the most extreme water depths and it means that the maximum damage already holds information about what part of the total value of the object or unit is susceptible to flood damage. The maximum damage does not include value which is not or unlikely to be susceptible to floods, such as the value of the land surface, the costs of building the foundation or the value present on high floors in buildings that are unlikely to collapse. Not all damage methods use this definition. Some include more items in the maximum damage and apply damage functions which never reach the value of one if part of that value is on average not susceptible to flooding. When comparing different models, the definitions of maximum damage and damage functions first need to be aligned to make a fair comparison.

In this chapter we discuss flood damage models, which we define as a set of maximum damages, damage functions, object data and their relationships with which a damage estimate for a flooding in a certain area can be made.

2.2.2 Types of uncertainty

In the uncertainty analysis in flood damage assessment two types of uncertainty are distinguished: Aleatory and epistemic uncertainty (Merz et al., 2009).

Aleatory uncertainty is related to the variability or heterogeneity within a population which can be expressed by statistic parameters such as the mean, variance, and skewness. This uncertainty is introduced by using average data: we use the maximum damage value of an average residence, although we know that some houses will suffer more, and other will suffer less damage. In small flood events which only affect a few houses, these few houses may differ significantly from the 'average house' and therefore the damage estimate for these houses is uncertain. In large flood events which affect many houses it is likely that deviations from the mean damage cancel out. This means that for large floods this type of uncertainty is of lesser importance.

Aleatory uncertainty by using averages can sometimes be reduced by applying more differentiation. E.g. the uncertainty within the maximum damage of a residence is reduced by using more differentiation in house types. The variation in maximum damage per house type is then less than if all houses together would be considered as one category 'houses'.

Epistemic uncertainty is the lack of understanding of a system and can in theory be reduced by further study or by collecting more or better data. In other words, also the average damage itself is not certain. For flood damage assessments, data is only available for a small number of events and those events often differ significantly from each other. This variation between events is still poorly understood and is therefore related to epistemic uncertainty. The epistemic uncertainty as stated above is not reduced when many objects are flooded. Therefore, it is the dominant uncertainty type for large flood events.

This type of uncertainty is especially relevant when a damage module developed for one area is applied to another area. In such a case, for example the maximum damage values within the model related to houses, may not be valid for the types of houses in the area under consideration. It would therefore be good not to mix up maximum damages and damage functions from different areas. However, given the scarcity of data and flood damage models, this leaves many modelers with the difficult choice between damage functions and maximum damages based on recorded data from another area or a local estimate.

2.2.3 Uncertainty in unit loss method

Uncertainty in the unit loss method consists of uncertainty in object data, in maximum damage figures and in the damage functions. Table 2.1 shows an overview of the uncertainties present in these aspects of the unit loss method.

Table 2.1. Overview of the uncertainties in flood damage modeling.

Element	***Uncertainty***	***Type***	***Expected significance***	***Included in analysis***
Object data	Quantity	Both	Depends on input data, expected to be often insignificant.	No
	Location	Both	Depends on area, often insignificant	No
Maximum damage	Value of the object	Mostly aleatory	Varies	Yes
	Susceptible to flood damage	Mostly epistemic	Significant	Yes
Damage function	Parameter representation	Both	Significant	Yes
	Knowledge/data about damage	Epistemic	Significant	Yes

2.2.3.1 Uncertainty in object/land use data

Uncertainties are found in the quantity of objects and their precise location. The precise location of objects is important, since the flood hazard characteristics (e.g. depth, or flow speed) may differ substantially from location to location. Each object should be linked to the hazard value present at the location of the object. The effect of the uncertainty on the precise location of objects on damage estimates is smaller for more homogenous hazards. For example, in a deep flat polder the exact location of an object is not important because the water depth is approximately the same everywhere. When a hazard becomes more heterogeneous the uncertainty in the exact location becomes more relevant.

Geographical location data uncertainties are especially significant in areas which flood frequently, but with small water depths. Because, in this type of area a small error in the location or height of an object can make the difference between an object getting wet frequently or very rarely. Furthermore, valuable objects susceptible to flood damage are unlikely to be placed on a location that floods frequently, so it is much more likely to count too many flooded objects than to count too little. Therefore, damage estimates may be very wrong if the approach used was too coarse to see local elevations or the exact locations of objects. For example, the Dutch standard damage model HIS-SSM estimated 100 million euro damage for an event in an unprotected area that in reality had only caused about 30 000 euro in damages (price level 2012) (Slager et al., 2013).

Such errors are, however, unlikely when objects are not elevated on purpose, placed on safe locations or protected in other ways. Without this local protection for some objects the damage will be overestimated and for others it will be underestimated. If many objects are affected, these errors compensate each other which reduces the uncertainty in the total damage. Use of high resolution elevation information is also very useful to reduce this uncertainty (Koivumaki et al., 2010).

Uncertainty in the quantity of objects can be caused by errors in data, or by using data sources that are inappropriate for the intended application. This uncertainty depends on the quality of the dataset that is used. De Moel and Aerts (2011) illustrated that this type of uncertainty may be small as they showed that different types of land use maps for the same area only have a small impact on the resulting damage estimate. In the uncertainty quantification for this paper the uncertainty in the geographical location data is neglected.

Also objects can be represented in different ways and each way can cause different uncertainties. A company office for example, can be represented by either: The floor space in the office, the footprint area of the office building, an area larger than the office building on a rough land use map or the number of jobs within the office. All these object representations correlate in some way with value present in the building but the indicator will not precisely correspond with the value of the building. The uncertainty this causes is aleatory, because if the indicator overestimates the value at one point it will underestimate it somewhere else, assuming that the maximum damage is a good average.

2.2.3.2 Uncertainty in maximum damage figure

The uncertainty in the maximum damage figure can be divided into two parts: The uncertainty in the value of the object and in the part of that value that is susceptible to flood damage.

There are generally two ways to obtain the maximum damage for a flood damage model: Deriving this from economic data or by looking at synthetic (hypothetical) buildings.

Economic data typically provides a total value per sector of all physical assets in the economy. To obtain a maximum damage figure per unit, this total value can be divided by the number of units within that sector. Next, the part of this total value which is susceptible to flooding must be identified. The strength of this method is that the mean object value will be accurate. However, uncertainty is still present in the part of the object that is susceptible to flood damage. A similar method is to use average construction costs and to correct this for the fraction that is actually susceptible to flooding.

Alternatively, the maximum damage of a category can be obtained by defining a “hypothetical” average company or object, and assessing the damage of all parts/aspects within that hypothetical company. The strength of this method is that the damage function and the maximum damage are well connected. Furthermore, the part of the value that is susceptible to flood damage is determined in a systematic way. The disadvantage of this approach is that epistemic uncertainty is introduced in the value of the object as the assumptions about this may be wrong.

Because generally a lot of good data is available about the value of objects, the uncertainty in the value of objects is considered aleatory. The uncertainty in the part of the value which is susceptibility to flooding is epistemic, because little data or knowledge about that is available.

2.2.3.3 Uncertainty in damage functions

Damage functions can be obtained in two ways: by analyzing data on observed damages to objects and flood characteristics in past flood events, or by defining hypothetical 'average' objects and assessing their damage corresponding with different flood intensities. Also a combination of both approaches may be used.

Flood damage data is rarely collected in a systematic way (Thieken et al., 2005), and not always available for research. When it is available it is often limited to a single or a few events. These events are often not representative for other types of floods or other countries or areas. Cultural or geographical differences can cause the use of different building or interior materials between regions and events, making one dataset not applicable to other areas. Another problem is that data is often limited to certain ranges of a flood parameter. For example, data may be only available for low water depths or the flood that was the source of the data may have coincided with a storm. In such cases the data cannot be used for events with larger water depths or no storm.

In general, transferring data from one event to another is error-prone. This makes it very difficult to apply knowledge derived from one event on another. Even when the data is applied to the same area as the data was taken from, problems may arise. Different flood events in the same area may lead to very different damages due to different human responses. For example, the same area in the Netherlands flooded in 1993 and 1995 with approximately the same water levels. The second time the damage to housing content was about 80% less (Wind et al., 1999). Also the damages due to Rhine floods of 1995 were less than half of the damages that occurred in 1993, as a result of precaution measures taken by households (Bubeck et al., 2012). This shows the sensitivity of flood damage to other factors than water depth. These other factors (in this case flood experience) are often neglected in the recordings. This example shows that a dataset based on a small number of events does not capture all possible variable values.

Synthetic damage functions solve many of the problems of having too few empirical data on actual damages. In this method a hypothetical building is defined and flood damage is assessed for each building part. The hypothetical building should be representative for the average building in the area. When it is not, or when the damage estimates for the different building parts are not right, the damage function is inaccurate.

Damage data can also be combined with expert knowledge. Probably the most common method to create a damage model is by picking and choosing damage functions from other models based on an analysis of which existing damage function best represents the area considered. Or, the average between different functions could be used as a damage function. The challenge with this combined method is to understand the background assumptions between the models that are brought together or compared. For example, a common challenge may be that the maximum damage definitions don't match.

The ideal case is to combine the best of the two methods. The damage data available should be used to calibrate a synthetic model. This limits the possibility that large errors are made in the interpretation of the damage data (e.g. wrong definition of the maximum damage), by forcing the modeler to think about the processes. Furthermore, it gives the modeler the freedom to diverge from the observed data in situations that don't match any of the recorded events.

A common problem in constructing damage functions is that it is difficult to include the large number of parameters that may influence the flood damage. The parameters that are not used are implicitly considered. Each flood damage model based on a limited number of parameters is therefore making assumptions on the effect of the non-explicitly considered parameters. Those non-considered parameters have been very significant in a subset of flood events. For example, in the 2002 Elbe floods contamination was critical (Thieken and Muller, 2005), in the Meuse floods flood experience was critical (Wind et al., 1999) and in the 1945 floods in the Wieringermeerpolder in the Netherlands the waves in the flood water were critical (Duiser, 1982). This last example is complicated by a study of Roos (2003) who showed that the findings of the 1945 Wieringermeerpolder flood are not valid for modern buildings. So also the construction year/type of a building can in some cases be a critical parameter. Other possibly significant parameters are, for example: Building style, flow velocity, flood duration, warning time and preparation.

Parameters that are not used can have a correlation with parameters that are used. For example, the water depth is correlated with the flood duration for floods in the Netherlands (Duiser, 1982; Wagenaar, 2012). Because of this correlation, the uncertainty caused by not knowing the flood duration is limited in the Netherlands. This relationship between two parameters may, however, be completely different for other types of floods (e.g. flash floods). A generally applicable flood damage model therefore still needs all parameters.

Table 2.1 splits the uncertainties involved in flood damage functions into two groups: Those related to parameter representation (using fewer parameters than theoretically necessary to describe the damage processes) and those related to lack of knowledge about the damage processes. The lack of knowledge related uncertainties are epistemic while the group of uncertainties related to the use of fewer parameters than necessary is are aleatory, because this group of uncertainties would remain even with perfect knowledge. In sub-chapter 2.3 the analysis of the epistemic and aleatory uncertainty components is used to assess the uncertainty in the outcome of a single damage calculation. The epistemic uncertainty will be estimated by using the difference between damage functions from entirely different flood damage models (section 3.1.2). The aleatory uncertainty will be considered by looking at the variation within one flood damage model (e.g. the difference between the low and the high estimate of the same flood damage model).

2.3 Methodology

2.3.1 Overview of the method

This chapter proposes a method for quantitative uncertainty analyses using a Monte Carlo analysis.. The qualitative uncertainty analysis discussed in the previous section is used to estimate the uncertainty in the inputs and the correlations between the different input parameters. The general assumption behind this uncertainty analysis is that no good local damage functions are available and that the modeler therefore doesn't know what damage functions to choose. For areas for which good local damage functions are available the approach discussed in this chapter may overestimate the uncertainty.

The damage analysis in this paper is limited to two damage categories: Houses and companies; as they are represented in the different flood damage models on which this research is based. These were selected since all damage models contain functions to assess them. This is not the case for other damage types. In flood damage models that do consider other damage categories, houses and companies usually make up the majority of the direct damage. Both damage categories are divided into damage to buildings and damage to content. Many individual flood damage models provide several more detailed sub categories for these basic categories. Our approach may therefore lead to slightly larger uncertainties than present in such more detailed models.

A crucial aspect of the Monte Carlo analysis of uncertainties is the correlation amongst the uncertainty of the different input parameters, such as the maximum damages of houses. If the maximum damage of for example house X is overestimated, the maximum damage of house Y may also be overestimated. The parameters that are homogenous within one event, but vary between events will have a strongly correlated uncertainty value: e.g. if damage depends on warning time and in one particular event the warning time is unusually short, this is more or less the case for all houses which were affected in that event. Such aspects are therefore sampled for the entire area at once. Other parameters vary between neighborhoods, or from place to place, such as for example the building type. Those need to be sampled on a smaller level then the entire area at once. Sampling will therefore be done at two different levels: For the entire event and on a more detailed sub-event level.

Figure 2.1 gives an overview of the calculations process which is repeated ten thousand times. This results in ten thousand different damage estimates which together make up the distribution of possible damages.

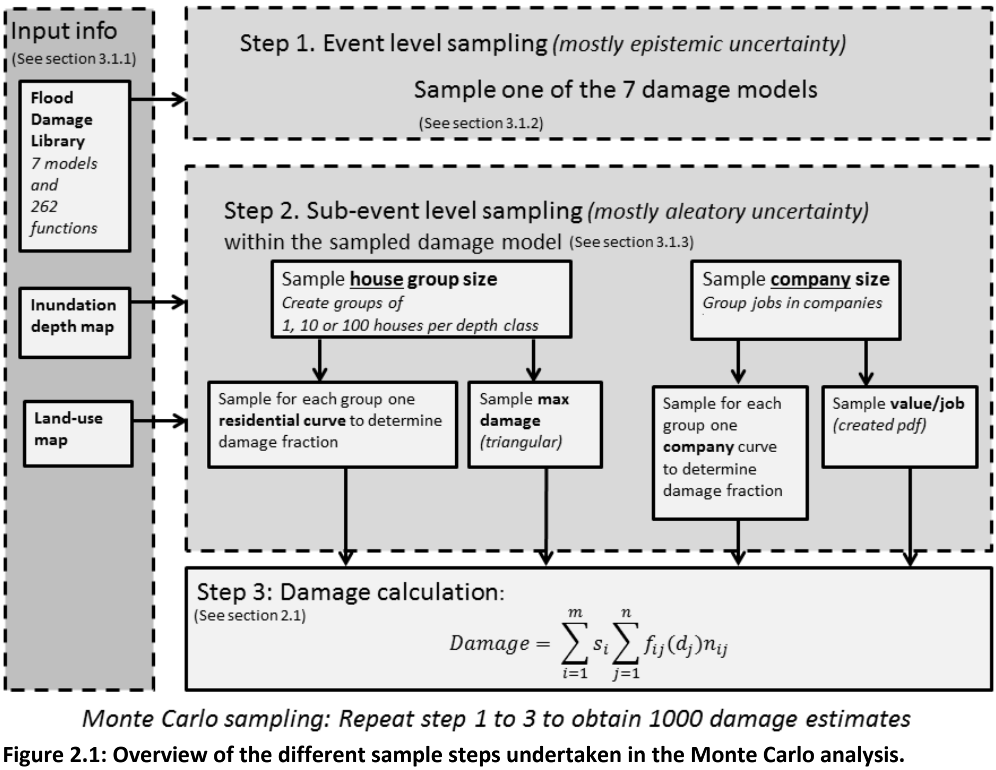


Figure 2.1: Overview of the different sample steps undertaken in the Monte Carlo analysis.

2.3.1.1. Input information

Flood damage library

A damage function library was constructed containing 262 different damage functions from 7 different flood damage models. Damage functions from flood damage models were included if they were made for developed countries and available to the author at the time of the study (2013-2014). These functions were the basis for the damage fraction and the susceptibility to flooding. The damage fraction was sampled by picking damage functions from a flood damage model. These functions were individually all scaled to one to ensure that the same maximum damage definition is applied everywhere.

Table 2.2 gives an overview of the models included in the damage function library. The Tebodin model only has damage functions for companies and the Billah (2007) model only has damage functions for houses. Since both models were made for the Netherlands and the damage functions were constructed using similar techniques, these two models have therefore been merged into one flood damage model.

Table 2.2. Overview of the damage models from which damage functions are included in the damage function library.

Model	Description
HIS-SSM	The standard Dutch flood damage model (Kok et al., 2005). It is based on several earlier Dutch flood damage studies (Duiser, 1982; Briene et al., 2002). The functions are based on expert estimates combined with data from the 1953 flood in Zeeland and the 1945 flood of the Wieringermeerpolder.
HAZUS-MH	An American disaster impact model with a flood module. This model was created by the federal government agency FEMA. It is described in FEMA (2008) and Scrawthorn et al. (2006). HAZUS provides a large set of American flood damage functions. From the HAZUS library a subset was used in the library presented here. The functions taken for houses were based on American insurance data and the functions for companies are based on expert judgment from the USACE.

MCM	The Multi Coloured Manual (MCM) is a British flood damage model. For the library presented in this document the version of Penning-Roswell (2005) was used. MCM-based on a systematic expert judgment approach were a hypothetical building is split up in smaller parts, with each part being evaluated separately. The model has a large number of functions for different types of company buildings.
FLEMO	A German flood damage model based on data from the Elbe floods of 2002. The functions were derived from FLEMOps for houses (Thieken et al., 2008) and FLEMOcs for companies (Kreibich et al., 2010). The functions include a low and a high estimate.
Rhine Atlas	This second German model is based on expert judgement taking into account and data from an earlier German damage database (HOWAS). More information about these functions is available in Jongman et al. (2012).
Tebodin	This is a Dutch study, based on a detailed, systematic and well documented expert judgement approach. This study only provides damage functions for industry. It is detailed: it provides functions for many different industrial types and it has separate functions for areas protected by flood defences and for unprotected areas (Snuverink et al., 1998 Sluijs et al., 2000).
Billah, 2007	This is a research project in which the systematic expert judgment approach as used in MCM was applied to Dutch houses.

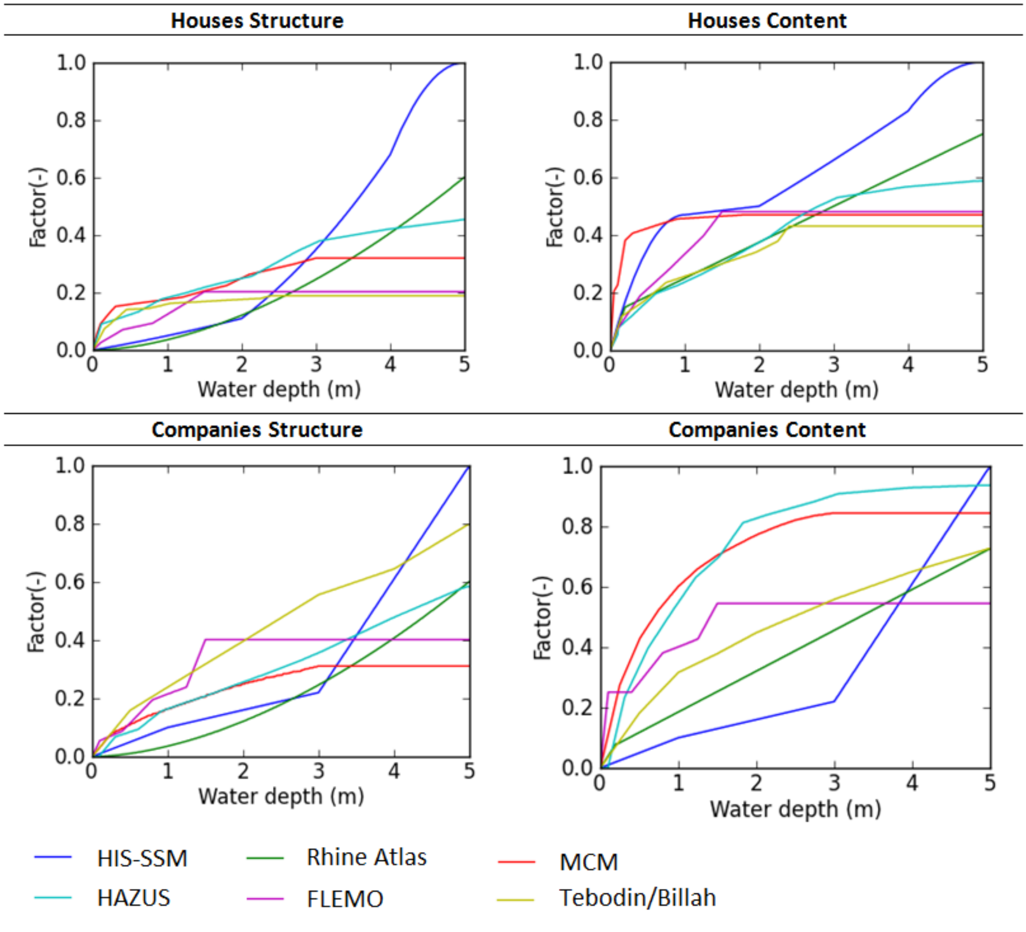


Figure 2.2: Average damage functions for the different flood damage models in the damage function library. Blue: HIS-SSM, Green: Rhine Atlas, Red: MCM, Light blue: Flemo, Pink: HAZUS, Yellow: Billah(2007) for houses and Tebodin for companies.

In this library the definition of the zero height point is the ground level (elevation in digital elevation model) rather than ground floor level. Some flood damage models use the floor level as the zero point and combine this with vertical elevation data of the ground floor. For this paper no damage functions with significant damage below the zero point were used. In the few functions used with damage below the zero point the zero point was shifted to the point that the first damage occurred on to make them comparable with other functions.

Land-use maps

A flood damage model needs input about the number of houses and jobs affected. For the case study the number of houses was taken from the geographical database BAG. This database is made by the Dutch Cadaster, Land Registry and Mapping Agency. For the number of jobs the background data of HIS-SSM was used (Kok et al., 2005).

2.3.1.2. Step 1: Event-level sampling (epistemic uncertainty)

The sampling on the event level is done by sampling a flood damage model (e.g. HAZUS or MCM) and using that throughout that damage calculation. This sampled model will be applied to all categories and will be used as source for the damage functions and for the susceptibility to flooding of the maximum damages. The advantage of this is that a realistic combination of inputs will be sampled. This procedure prevents that on average a higher damage for small water depths than for large water depths are sampled or that functions with different implicit assumptions are merged.

3.1.3. Step 2: Sub-event level sampling (aleatory uncertainty)

Group size and dependency

For the sub-event level sampling, uncertainty values are sampled for small groups of houses or for a company. Houses and jobs are grouped because in reality also often similar houses are built near each other and also a company is expected to be relatively homogenous in damage per job. It is therefore not realistic to sample all houses and all individual jobs in an area completely independent from each other. By sampling in small groups of houses/jobs total dependency is assumed within the group and total independency between the groups.

The way in which the area is grouped determines the dependency for this aleatory uncertainty. This buildup of groups is therefore also sampled again for each Monte Carlo simulation. This sampling should therefore be seen as a sampling of the dependencies between the damage of different objects. For houses the area is split in groups of 1, 10 or 100 houses for every simulation. Houses are only grouped when they have a similar water depth. This is done to keep the calculation simple but also because similar water depths typically occur in locations that are geographical close. Furthermore, the group sizes are so small that for medium, or larger, sized events this assumption has no influence on the results. For companies, the jobs are grouped per company.

Damage functions

Within each group every house/job receives the same damage fraction and maximum object value. Sampling for the damage fraction is done based on the set of damage functions within the flood damage model sampled in step 1. For example, if the flood damage model sampled has 3 damage functions for houses, for each group one of the three damage functions is randomly used.

Maximum damages

The values are based on De Bruijn et al. (2014) and Gauderis (2012). De Bruijn et al. (2014) estimated the structural value of a **house** on € 125 000. The minimum and maximum from the triangular distribution are roughly estimated at +/- € 75 000 for structural damage. For content damage De Bruijn et al. (2014) estimate a maximum damage of € 70 000, for which here also a triangular distribution is assumed, with +/- € 50 000. These maximum damages all use the price level 2011. These assumptions lead to a symmetric probability distribution, while it is probably in reality positively skewed. This is neglected in this study because it is difficult to estimate and the impact on the uncertainty is expected to be very small.

The maximum damage for companies is in this study assessed as the maximum damage per job times the number of jobs per company. Gauderis (2012) estimated material value per job for 62 different categories of **companies**. These estimates are taken together to produce a distribution of the physical value of a company per job. Because not all company categories are equally common, the values were weighted in the distribution based on their quantity in the Netherlands. This more complex method was used in order to correctly incorporate the skewness and because the data from Gauderis (2012) makes it possible to do this, which was not the case for houses. The results are shown in figure 2.3. These values include both the structure and the content. Assumptions from Gauderis (2012) on the part of the maximum damage that belongs to the structure and the part that belongs to the content were adopted.

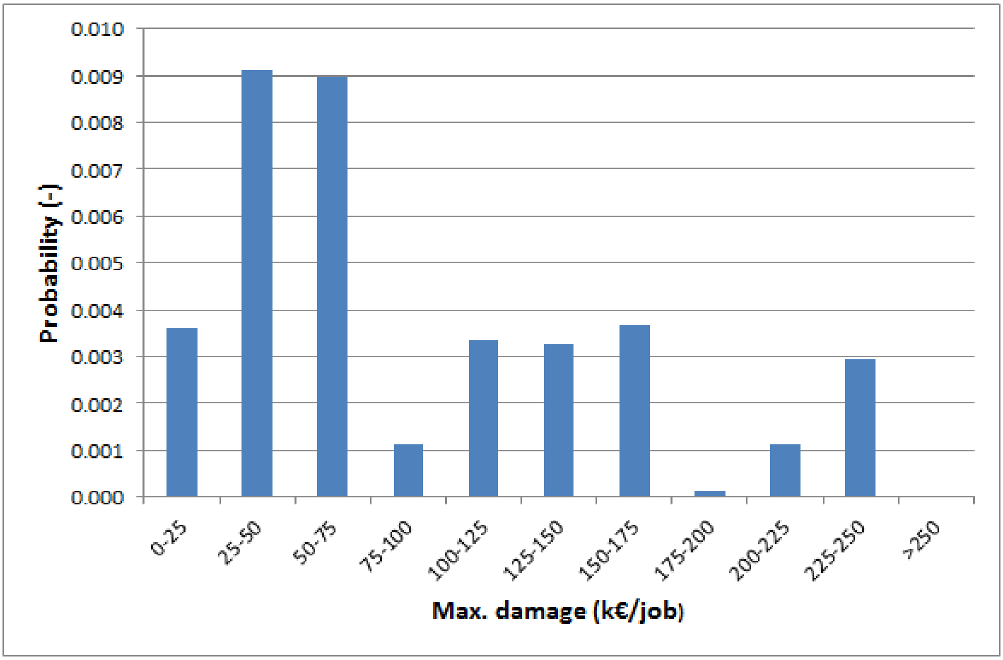


Figure 2.3: Probability on different maximum damages as it was used in the calculation. This is based on a study of Gauderis (2012) who calculated the damage for 62 different company types in the Netherlands. Data about how frequently these company types occur in the Netherlands is used to make this estimate of the probability density of the maximum damage per job in the Netherlands.

2.4 Monte Carlo analysis behavior: Hypothetical flood maps

To gain understanding of the Monte Carlo analysis behavior it was tried on hypothetical flood depth maps, one with small water depths (<0.5m), one with medium water depths (0.5 – 2m) and one with large water depths (2-3m). These had average water depths of 0.35, 1.25m and 2.5m. These maps were used for calculations with 150 and 15000 houses and jobs. In total thus 6 different trials were carried out and the resulting uncertainty values were compared.

The uncertainties in the damage estimates are expressed with the coefficient of variation. This is the standard deviation of the damage divided by the mean of the damage. It has no unit and is therefore independent of the size of the flood event. This makes it a good measure to compare the uncertainties in different areas.

Figure 2.4 shows the results of this hypothetical analysis. It stands out that both a smaller water depth and a smaller area increase the uncertainty significantly. This is because at small water depths the different flood damage models differ significantly more from each other than at large water depths. This indicates that the uncertainty in damage estimates for events like for example small regional levee failures is much larger than the uncertainty in damage estimates for large scale floods with large water depths.

Another observation is that the distribution of the damage for small events looks very different from the distributions of the damage of large events. The main reason for this is that for large events the aleatory uncertainty in the flood damages can be reduced significantly by the law of large numbers, but not the epistemic uncertainty.

Epistemic uncertainty is therefore the significant uncertainty for larger events. The frequency distributions therefore show then clearly separate peaks related to the damage functions of the separate flood damage models.

It is difficult to determine for what event size the variation between the flood damage models (epistemic uncertainty) becomes more important than the variation within the flood damage models (aleatory uncertainty). For the uncertainty model created in this paper this point is somewhere between 100 and 3000 houses plus jobs. This critical size depends on the dependencies between individual objects. These dependencies determine how fast the law of large numbers will reduce the aleatory uncertainty. For this paper this was estimated by sampling in groups instead of in individual objects. The size of these groups therefore determines when the epistemic uncertainty becomes dominant. These group sizes were based on a rough estimate in this paper and more research should be done for better results.

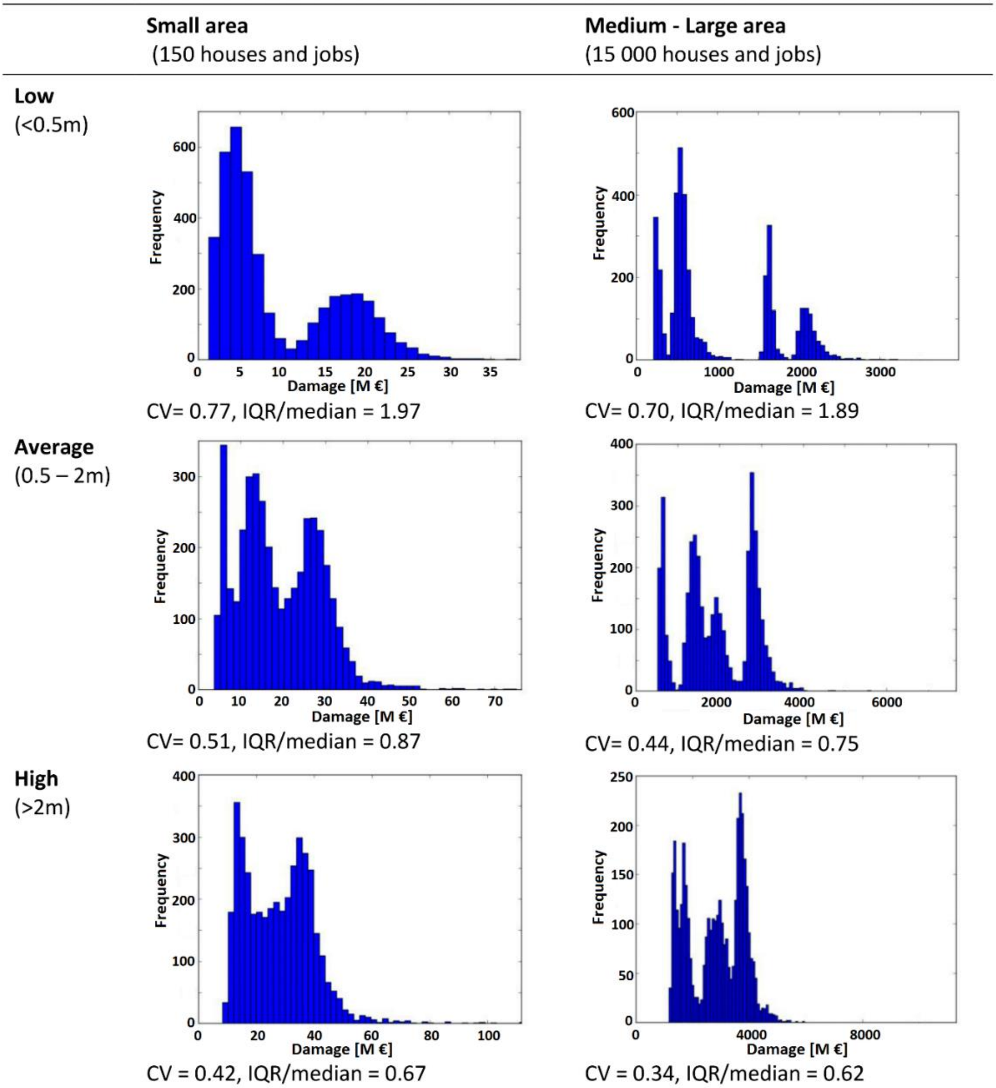


Figure 2.4: Results of the Monte Carlo simulations applied on synthetic flood maps shown as frequency distributions, coefficient of variations (CV) and IQR (interquartile range)/median for different types of hypothetical areas based on 4 000 samples. The CV and IQR/median are both simplified (imperfect) representations of the dispersion in a single number, they do show a similar view of the dispersion for the different test runs. The x-axis in this figure is equal to four times the mean damage.

2.5 Case study

A case study is done in the Betuwe, Tieler-en Culemborgerwaarden area (dike ring 43) in the Netherlands to show the effect of the uncertainty in the flood damage estimation on investment decisions for flood risk management. Dike ring 43 is located between Rhine branches in the Netherlands. In the west the area is closed with a high dike (border to next dike ring area). The area slopes down to the west. The difference in height between the eastern and western part is about 10 meters.

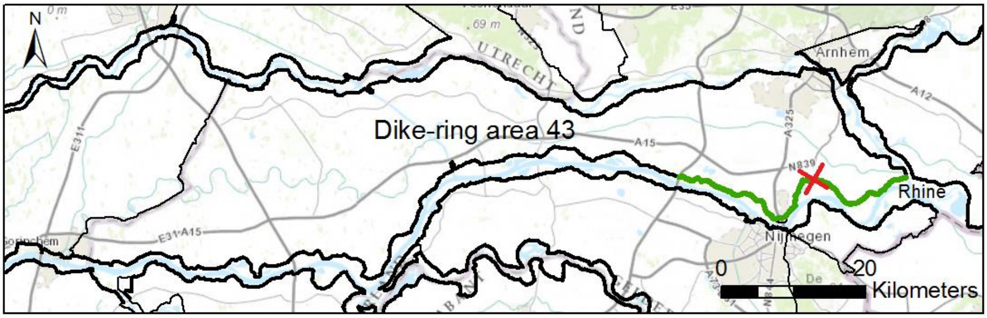


Figure 2.5. Map of the case-study area. The green line is the dike segment that is looked at, the cross indicates the location of the breach scenario which was used.

The Monte Carlo analysis is applied on a water depth map resulting from a simulated dike breach (VNK, 2014) along the Rhine river near Bemmel in the Netherlands (see figure 2.5 for its location). This dike section is about 26 kilometer long. Bemmel is situated in the eastern upstream part of the Betuwe area. When the dike breaches, water flows through the Betuwe area to the west where it is stopped after about 70 kilometers by the western embankment. The maximum water depths due to this dike breach vary from less than 50 cm in the east to over 5 meters in the west. In this dike-breach scenario a total area of 626 km^2 is inundated. This area contains several small cities and villages, with a total population of around 300 000 people. The large flood extent and the large number of affected residences and companies and the large variation in water depths are expected to have a reducing effect on the aleatory uncertainty in the total damage of the dike-ring area.

The damage assessed for this flood scenario was 16 billion euro (price level 2011) with a standard deviation of 5.6 billion euro based on 10000 simulations. The resulting damage outcomes are shown in figure 2.6. The peaks in figure 2.6 are related to the damage models and illustrates the large differences between the different damage models (two damage models overlap).

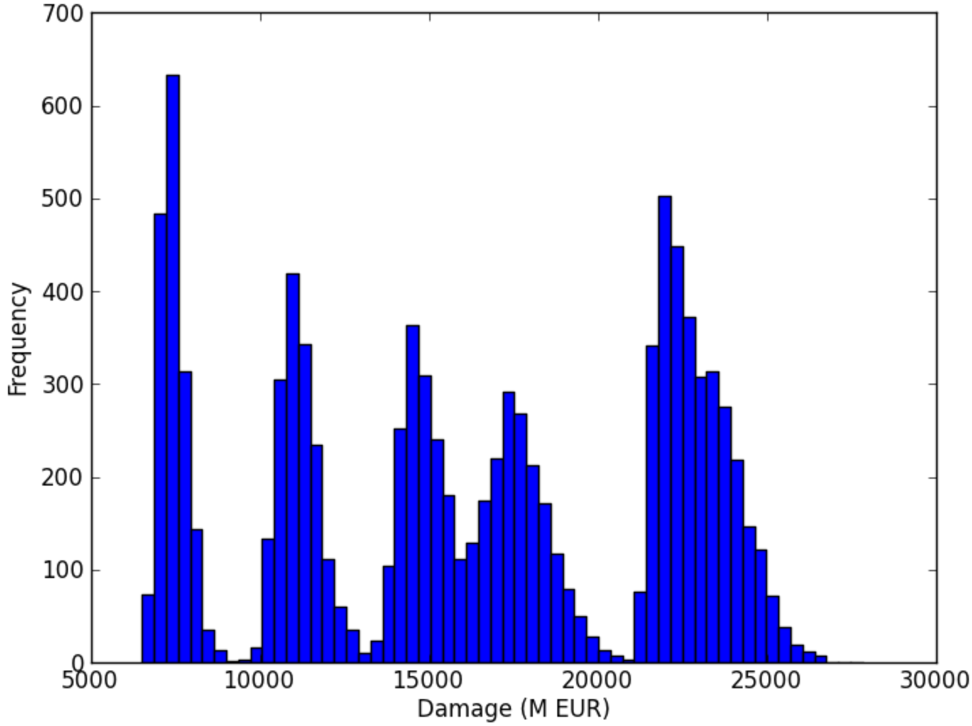


Figure 2.6: Frequency distribution of the damage in the case study area (based on 10,000 Monte Carlo simulations). The 5 peaks represent the 6 models used in this study, with two models overlapping. The damage has the price level of 2011.

The results in figure 2.6 are used to find the economic optimum flood protection standard and investment strategy for the dike segment from an economic viewpoint. The optimum flood protection standard and investment strategy is calculated using a simplified version of the approach of Kind (2013). Kind (2013) assesses which investment strategy (set of dike improvements at different moments) has the smallest total costs for the future and hence is the economic optimum. These total costs consist of all the discounted expected annual damages (EAD) considering future changes and the discounted future investments in the dike. The expected annual damage depends on the flood probability and the flood damage given a dike breach (as calculated in this case study). Long term economic growth forecasts are used to increase the damage for every year into the future. Furthermore, a correction factor was used to take into account indirect damage. The flood probability depends on the quality of the dike which again depends on the investments carried out. This flood probability is calculated for each year into the future based on the current status of the dike, investments up to that point, consolidation of the dike and climate change predictions. Investments in the dike are simplified as height increases and the cost of these investments is based on fixed and variable cost for the dike segment considered. A height increase is converted into a flood probability reduction based on a parameter that describes the height increase necessary to decrease the flood probability by a factor 10. This parameter and all other parameters used for determining the optimal investment strategy (except for the flood damage) were taken from the WV21 project (Kind, 2011).

In this paper we assess the effects of uncertainty in damage estimates on the economic optimal flood protection standard and the total investment costs. We do that by determining the investment strategy for five different damage estimates. The first four estimates relate to the first four peaks. For the highest damage estimate the 98% percentile of the damage outcomes was used.

The analysis in this paper focuses on the first investment made. In all five alternatives this investment is done in 2015. The second investment is in all alternatives planned about 75 years later and a third investment is suggested about 50 years after the second one (around the end of the time span considered). The total investment costs are mainly determined by the first investment, because the weight of later investments is very small due to the use of the net present value which gives future costs and benefits a much lower weight than current costs and benefits. The calculations assumed a discount rate of 5.5% (based on WV21 Kind (2011)).

Table 2.3: Optimal investment strategy given different damage estimates. The flood protection standard is a return period based on the direct method described in Kind (2011), note that the actual return period of the investment strategy differs per year. Price level 2011 is used.

Damage (M€)	***Extra height for the investment (cm)***	***Flood protection standard (y)***	***Present value Investment cost (M€)***	***Present value EAD (M€)***	***Present value Total cost (investment cost + EAD) (M€)***
8 000	70	25 000	109	11	121
12 000	77	40 000	114	11	126
15 000	80	50 000	117	11	128
18 000	82	58 000	118	12	130
25 000	88	83 000	122	12	134

The results in table 3 show that the optimal investment strategy is at first glance not very sensitive to the precise damage estimate. The difference between the five alternatives in required dike heightening is only 18 centimeters (88 minus 70 cm). This small difference is partly explained by the strong sensitivity of the flood probability to the precise height of the dike. The dike segment in this case study becomes 10 times safer by raising it with only 34 cm. If the flood probability would be less sensitive to height changes the differences in dike height between a low and a high damage estimate could be much larger. If the dike should be increased with 1 meter to reduce the flood probability with a factor 10, the difference between the top and lower damage estimate would be 47 cm.

If the flood damage applied in the cost benefit analysis differs from the flood damage that would actually occur, a suboptimal investment strategy would be applied. Table 2.4 shows the costs of using a wrong damage estimate. It gives the unneeded cost made by assuming a certain damage for different ‘real damage values’. This cost varies in this case study between 0 and 12 million euro and is on average about 2 million euro, which is 1.4% of the total costs and for this case study about 75 k€/km.

The maximum error is 9% of the total costs and for this case study 500 k€/km (price level 2011).

Table 2.4: Cost of a damage estimate error for this dike segment in million euro (price level 2011).

		Damage estimate for calculation investment strategy				
		8 000	**12 000**	**15 000**	**18 000**	**25 000**
Damage Reality	**8 000**	0	1.1	2.1	2.7	5.5
	12 000	1.3	0	0.2	0.5	2.3
	15 000	3.2	0.3	0	0	1.2
	18 000	5.2	0.9	0.2	0	0.7
	25 000	12.2	3.8	1.8	0.9	0

This case study illustrates how the Monte Carlo analysis may be used to assess the uncertainty in damage assessments, and how the effect of this uncertainty on investment costs may be determined. In the case study here, the effect is small. However, if we take into account the fact that in The Netherlands we have about 3000 kilometers of embankments and that 12 million might be unnecessary spend per 26 kilometer, the total amount of money spend unnecessary may then be large. It is also likely that in cases with lower flood probability standards, or with smaller flood events the effects of this uncertainty are much larger.

A striking observation in the results of table 2.4 is that the costs of overestimating the damage are significantly lower than the costs of underestimating the damage. The difference in costs is on average a factor 2 (see table 2.4). This can be explained by the non-linear relationship between the flood probability reduction and the investment costs. The flood probability can be reduced a lot with a small extra investment, thus when too little is invested the EAD goes up faster than the investment costs go down. This implies that under uncertainty it would be economically efficient to add a safety factor to avoid investing too little.

2.6 Discussion

This paper discusses a new method for the quantification of uncertainties and applied this method in a case study. The case study is a good illustration of the method and its use, but the calculated uncertainty, the damage frequency distribution and the effect of uncertainty on investment decisions, may not be representative for all situations. First of all, because several damage determining aspects were neglected in the case study. The damage is assumed to consist only of damage to buildings and companies. Other damage categories, such as affected persons and fatalities may also be relevant to quantify and can be taken into account in a CBA (cost benefit analysis). Another simplification is that the entire cost benefit analysis in the case study is based on only one flood scenario at one breach location and at one water level (at the design water level of the dike). A more precise way would have been to include multiple breach locations and water levels. These effects are however assumed to be negligible for the conclusions of this paper, because in most cases the uncertainty for the different damage estimates will be similar and highly correlated (because its for the same area).

Secondly the results may not be representative for all situations, because the exact location and number of peaks in the damage frequency distribution depend on the input damage models in the uncertainty analysis. The set of 7 damage models used does not cover all possible damage models. If an extra flood damage model would have been added to the damage function library an entire new peak could appear. The frequency distributions of the outcomes must therefore be considered as an example of what a frequency distribution could look like and how far the peaks are approximately apart from each other. It is impossible to make a real frequency distribution because the major uncertainties are epistemic uncertainties. Epistemic uncertainties are by definition not understood and can therefore not be represented by a frequency distribution (Helton & Oberkampf, 2004). However, there are alternative concepts to describe uncertainty such as imprecise probabilities or Bayesian statistics that deal with this problem (Reichert & Omlin, 1997; Zadeh, 2005).

Thirdly, the costs of a wrong estimate which were estimated for the case as about 1% and at maximum 10% of the total costs may also be different for other cases. It depends amongst others on the costs required to reduce the failure probability with a factor 10, on the damage itself and on the uncertainty in the damage interaction (which will be larger for small areas and areas with little flood water depths).

Finally, the uncertainty in the damage estimate was, in this case study, directly linked to an error in the investment strategy. However, in the determination of the optimal investment strategy not only uncertainties in the damage estimate, but also in other components play an important role. Uncertainty in the costs of dike strengthening, in the discount rate, in the future economic growth, in the flood pattern and so on, all add to the uncertainty in the optimal investment strategy. These uncertainties may partly compensate each other, but can also aggregate each other. Their relative importance differs per case depending on local characteristics (De Moel et al., 2014).

We tried to combine information from different damage models to get a better quantification of uncertainties in damage outcomes. This can only be done when the damage models may all be applicable to the flood scenario which is being modeled. Whether flood models are equally applicable is sometimes difficult to establish. Metadata of the source of the damage models is not always available and sometimes information on the event on which the model is based, is also lacking. This makes it difficult to compare damage models and to understand why they have different estimates for the same flood patterns. Relevant metadata on parameters which may be obvious for a certain event, but vary from event to event are needed. Examples of such parameters are for example flood experience of the population, building style, flood duration, contamination of the flood water, etc.

Metadata for flood damage functions should give clear instructions about the type of events for which damage functions are applicable and for what events they are not. This could lead to a classification of different flood types with their own damage functions. This would first lead to a better transferability of damage functions and maximum damages and could eventually lead to generally applicable flood damage models.

Vogel et al., 2012 and Schröter et al., 2014 have also done detailed uncertainty analyses for flood damage assessment. The most obvious difference with these studies is that this paper is using a Monte Carlo approach while they use fundamentally different methods. However, the more interesting difference is that this paper assumes a situation where no good local data is available and that little is known about the expected conditions during the potential flood (apart from the maximum water depth). Therefore, this paper used relatively simple data from many different countries and flood types as input for the uncertainty analysis, while these other papers used relatively complex data from only Germany. The strength of this approach therefore is that it has a wider coverage of the spectrum of possible flood damage. The disadvantage of this approach is that it is not applicable when a good local flood damage model is available based on a lot of data.

2.7 Conclusion

Uncertainties in flood damage estimates can be large. This study showed uncertainties of an order of magnitude of 2 -5. This uncertainty is mainly caused by a lack of knowledge. Most flood damage models are based on data resulting from a small number of events. Because flooding can occur in many different ways (water depths, contamination, flow velocities, flood durations, etc.) and in many different types of areas (building types, flood experience local population) any model will miss considerable parts of the spectrum of possible options. Data from one event therefore is often not transferrable to other areas or events. Since only data representative for the event under consideration can be used, little data is available and hence large uncertainties are introduced in flood damage modeling.

This study introduced a method to quantify these uncertainties using a set of damage models which have all been applied in the past to river floods (not flash floods) or storm surges in developed countries. To quantify the uncertainty a distinction was made between epistemic and aleatory uncertainties. Epistemic uncertainties are introduced by a lack of knowledge about the spectrum of possible flood events and areas in which they could occur. The size of this spectrum was for this study estimated by using the difference between flood damage models. Aleatory uncertainties are introduced by local variations between objects and circumstances. These uncertainties were for this study estimated with the variations within different flood damage models.

These aleatory uncertainties are large for small flood events and much smaller for large flood events affecting many objects. Epistemic uncertainties are not smaller for large areas, since they are not related to deviations of single objects from the average object for which the damage functions were derived. Epistemic uncertainties can only be reduced with new knowledge. The resulting Monte Carlo analysis therefore shows larger uncertainties for small areas. However, at a certain event size the epistemic uncertainties become dominant.

These uncertainties in flood damage modeling can potentially have a significant effect on investment decisions. In this study a case study was carried out to calculate the economic optimal investment strategy for a dike segment. This case study showed that uncertainties in damage estimates can lead to sub-optimal investment decisions. In the worst case scenario (maximum error in damage estimate), the difference between the total costs (remaining risks and investment costs) may be as high as 500 k€ per km dike (price level 2011). The expected difference between the optimal and sub-optimal investment strategy was, however, significantly lower (75 k€ per km dike). These findings need to be verified with further research in other areas.

The paper provides a good first approach for uncertainty quantification in damage estimates and shows how this approach can be used to improve investment decisions. Further research including other areas and more flood events is recommended to develop the approach further.

Acknowledgements

We like to thank William Lehman, one anonymous reviewer and the editor Prof. Bruce D. Malamud for their constructive feedback and helpful comments which helped to improve this paper.

3. DEVELOPMENT OF MULTI-VARIABLE FLOOD DAMAGE MODELS

This chapter is based on the paper "Multi-variable flood damage modelling with limited data using supervised learning approaches" written together with Jurjen de Jong (Deltares) and Laurens Bouwer (Climate Service Center Germany). It is published in the journal Natural Hazards and Earth Science Systems (NHESS). The reference is: Wagenaar, D., de Jong, J., and Bouwer, L. M.: Multi-variable flood damage modelling with limited data using supervised learning approaches, Nat. Hazards Earth Syst. Sci., 17, 1683–1696, https://doi.org/10.5194/nhess-17-1683-2017, 2017.

Abstract

Flood damage assessment is usually done with damage curves only dependent on the water depth. Several recent studies have shown that supervised learning techniques applied to a multi-variable data set can produce significantly better flood damage estimates. However, creating and applying a multi-variable flood damage model requires an extensive data set, which is rarely available, and this is currently holding back the widespread application of these techniques. In this paper we enrich a data set of residential building and contents damage from the Meuse flood of 1993 in the Netherlands, to make it suitable for multi-variable flood damage assessment. Results from 2-D flood simulations are used to add information on flow velocity, flood duration and the return period to the data set, and cadastre data are used to add information on building characteristics. Next, several statistical approaches are used to create multi-variable flood damage models, including regression trees, bagging regression trees, random forest, and a Bayesian network. Validation on data points from a test set shows that the enriched data set in combination with the supervised learning techniques delivers about a 15% reduction in the mean absolute error, compared to a simple model only based on the water depth, despite several limitations of the enriched data set. We find that with our data set, the tree-based methods perform better than the Bayesian network.

3.1 Introduction

Decision making in flood risk management is increasingly based on studies that quantify the flood risk rather than only the flood hazard. Flood damage estimation is therefore becoming increasingly important (Merz et al, 2010).Flood risk assessment supports policy makers to decide which flood risk management measures are most efficient in reducing flood risks and how much investment is cost-efficient. With the European Union Floods Directive (EC, 2007) now fully in place, national flood risk assessments are being developed with the final aim to support flood risk management plans. In the Netherlands, such flood damage assessment has been used to derive the optimal protection standard for flood protection (Kind, 2013; van der Most, 2014), using the current Dutch standard method for damage modelling (Kok et al., 2005). Also for insurance applications, more precise estimates of flood damages are required.

Flood risk assessments require flood damage models. These models typically predict the damage as fraction of the potential damage, based on the water depth, and average building repair and replacement costs for different types of buildings (Messner et al., 2007; Jonkman et al., 2008). Similar approaches are also applied to other natural hazards, for example for landslides (Papathome-Köhle et al., 2015) and the software package HAZUS can be used for floods, earthquakes and hurricanes (Scawthorn et al., 2006). Alternative approaches to calculate flood risk do also exist, such as vulnerability indicators (Papathoma-Köhle, 2016).

Simple flood damage models often don't perform well, as shown by their validation (e.g. Jongman et al., 2012). This is because water depth alone cannot explain the full complexity of the flood damaging processes and several studies have only found low correlation coefficients (typically below 0.5) between the water depth and the flood damage (e.g. Merz et al., 2013, Pistrika and Jonkman, 2009). Furthermore, often no local data is available on flood damage and therefore a relationship between the water depth and damage either needs to be estimated or transferred from other areas (Wagenaar et al., 2016). This can cause errors as simple models hold many implicit assumptions that may not be valid for the situation the model is transferred to. For instance, Elmer et al. (2010) showed that an event with a low flood probability could not use the same damage function as a flood event with a high probability. These implicit assumptions cause large unexplained differences between flood damage functions (Wagenaar et al., 2016; Gerl et al., 2016). Transferability however could be improved, when a model describes more variations of the damaging process, and when more variables are included in the damage models (e.g. flood probability is explicitly part of the model). Similar problems are also present in the modelling of other natural hazards. For example, Fuchs et al. (2007) found that building materials are very important for debris flow damage modelling and that models can therefore not always be transferred in space and time.

Current approaches suffer from two main limitations: first, they rely on limited information and usually only take into account water depth as a predictor, and use a deterministic relation between water depth and some fraction of average maximum damages; secondly, they are deterministic in nature, while it has been shown that uncertainties in this approach are large, but generally not quantified e.g. in the Dutch standard method (Egorova et al., 2008). Some of the multi-variable methods are able to provide probability distributions, rather than deterministic estimates of damages.

Recently, multi-variable flood damage models have been created with a German dataset based on telephone interviews. Thieken et al.(2005) found that apart from the water depth also the contamination of the flood water and precautionary measures were important to estimate the flood damage. In Thieken et al. (2008) these extra variables were included in a simple multi-variable flood damage model as a surcharge. Using information from this same database, Merz et al. (2013) used regression and bagging trees and Vogel et al. (2014) used Bayesian Networks to predict the flood damage. Spekkers et al. (2014) applied regression trees to estimate pluvial flood damage. Van Oostegem et al.(2015) applied the Tobit estimation technique to a multi-dimensional dataset in Belgium to estimate pluvial flood damages. These multi-variable flood damage models have been shown to perform better than simple flood damage models in Schröter et al. (2014) (up to 25% reduction in mean absolute error, MAE), both tested on their own dataset and on datasets from other floods (Schröter et al., 2014). Also, some multi-variable approaches (Bayesian Networks, Bagging trees and Random Forests) generate probability distributions of estimated damages, and thus provide information on uncertainties of the estimates. Therefore, multi-variable flood damage models look like a promising approach to improve flood damage modelling.

The application of multi-variable flood damage models for flood risk management studies is still difficult because of the large data requirements. Running a multi-variable flood damage model for a new area requires for every object several variables on the flood hazard and building characteristics that are not yet typically collected. Also creating new multi-variable flood damage models is currently rarely done because they also require records of flood damages at building level.

More commonly available (although still rare) are simple datasets that hold records with the flood damage that occurred for each building with sometimes a few other variables (such as location or water depth). Such datasets may have been created for compensation purposes or to build simple flood damage models but may miss most of the desired variables. An example of such a dataset is the flood damage dataset collected after the Meuse flood of 1993 in the Netherlands which is used here. Previously this dataset has been described in Wind et al. (1999) and in more detail in WL Delft (1994).In this paper we will explore the use of supervised learning techniques to build flood damage models based on a dataset that is very different from the datasets used in previous studies (i.e. the German dataset applied by Merz et al. (2013) and Schröter et al. (2014)).) The dataset in this paper was collected by insurance experts directly after the flood for compensation purposes and covers all affected buildings. This is different from the German data which was collected a year after the flood for research purposes based on a sample of the affected buildings. The data is also different in that in the original study only a few variables were collected, in contrast for the German dataset all variables (except return period) were based on telephone interview answers. In this study several methods are applied to enrich the Meuse 1993 flood damage dataset with extra flood hazard and building characteristic variables. We will answer the question of whether this enriched dataset from a different source then previous studies is also suitable to build a multi-variable flood damage model. The expectation is that the multi-variable models perform better than a model based on a single variable (water depth) and that even data with limited quality will improve the results.

2D hydraulic simulations of the 1993 flood on the Meuse are used to enrich the dataset with additional flood characteristics. Cadastre data is used to enrich the Meuse dataset with extra building characteristics. Four different supervised learning techniques are then applied to this enriched dataset: a regression tree, bagging regression trees, random forest and a Bayesian network. A part of the dataset will be held back and will only be used for validation. This validation is then used to determine whether the enriched dataset combined with supervised learning techniques performs better than a traditional damage function based on the original dataset of water depths. In this paper we will focus on predicting absolute flood damages rather than relative flood damages. This is because the exact building values are not available.

3.2 Methods and data

3.2.1 Datasets

3.2.1.1 Meuse 1993 damage dataset

The dataset available for this research is based on the Meuse flood of 22 December 1993 in the Province of Limburg in the Netherlands (WL Delft, 1994). Although no dike breaches occurred in this event, several towns and urban areas located close to the river were affected. The flood caused a total of 254 million guilder (price level 1993) in direct damages, which is approximately 180 million euros today (price level 2016). The flood inundated 180 km^2, which is about 8% of the Province of Limburg. 32% of the damage pertains to residential buildings and content (furnishings). In this study only residential damage is considered. Other major damage categories were business (29%), government (24%) and agriculture (8%) (WL Delft, 1994). These categories are not considered because they are more heterogeneous and less data about them is available.

Damage information was collected in the context of a compensation arrangement for flood damages by the national government. All data was collected by sending damage experts from insurance companies to the affected buildings, several weeks after the flood event had occurred. Directly after the damage data was collected in 1994, the data was shared with WL Delft (now Deltares) to create a flood damage model. WL Delft received 5780 records for damage to residential buildings, 1382 records were incomplete and hence only 4398 records could be used. The damage to privately owned residential buildings was collected by an organisation called "Stichting Watersnood 1993" , the damage to companies and the structure of rental residential buildings was collected by another organisation called "Stichting Watersnood Bedrijven 1993". So, in this set up of the damage collection, the building structure of rental residential buildings was collected by "Stichting Watersnood bedrijven", the organization that collected company damages. This is different from the organization that collected the rest of the residential damages. The structure damage to rental residential buildings was only shared with WL Delft (1994) in some partial aggregate form. WL Delft (1994) presumably distributed this partially aggregated rental residential building damage over the individual rental residential buildings. The exact method for this was however not reported and the original dataset is no longer available. Therefore, we had to work with a dataset which includes unknown manual actions. The structure damage data is therefore from inconsistent quality, the content damage however has no such problems.

Furthermore, it is expected that the percentage of rental residential buildings in the affected area of Limburg is relatively low, limiting the impact of this data problem.

Another issue with the dataset is that for privacy reasons the exact locations of the buildings were not shared with WL Delft. Only the 6 digit postal code was available for this study, which makes it difficult to enrich the dataset, as between 1 and 20 buildings share the same 6 digit postal codes in the dataset.

In the original dataset the water depth (relative to the ground floor level) was estimated by the experts that surveyed the damage. The quality of the water depth estimate is questioned by WL Delft (1994; report 9, appendix A) because it was not the main aim of the survey and the experts visited several weeks after the water had receded. A plot of the water depth (see figure 3.1) and the damage doesn't show an obvious relation. The correlation between the water depth and the damage is weak (Pearson correlation coefficient = 0.18).

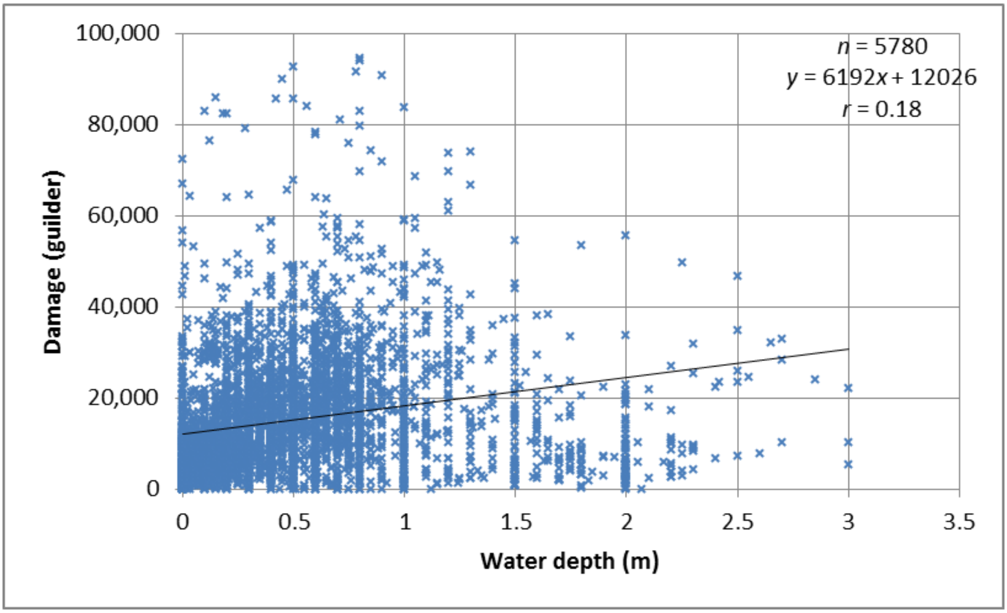


Figure 3.1: Scatter plot showing the relation between water depth and damage in the original data set..

The final dataset also contains information on the number of inhabitants per building, whether the house has a basement and whether the house was attached to other houses. However, this data is not described in any of the available reports so the collection methods are not known, but the recorded values are clear enough to incorporate in this study. Two more variables are also included in the WL Delft dataset and also not described in any available report. These are emergency actions and ownership of the house. The meaning of the values found in the dataset for these variables is however not sufficiently clear, and could unfortunately not be taken into account in this study.

3.2.1.2 Upgraded Meuse 1993 dataset

To improve the dataset, additional information is required on both the flood hazard and exposure variables. The results of a 2D flood simulation and cadastre data were used to upgrade the dataset, in terms of hazard and exposure information, respectively. Because no observational data is available on flood characteristics other than the water depth, a simulation of the flood event was done. In the 2D flood simulation tool WAQUA (Rijkswaterstaat, 2013), a verified model of the state of the Meuse during the 1993 flood was available (Becker, 2012) and applied in this study to get extra variables. Using this model, a new simulation was run using a discharge boundary condition at Eijsden and a water level boundary condition at Keizersveer for the period 1 November 1993 to 31 Januari 1994. This simulation was used to create a maximum water depth map, a flood duration map, a flood return period, and a flow velocity map at a spatial resolution varying between 10 and 40 meters.

The maximum water depth and flow velocity are standard outputs of WAQUA. Flood duration is however not a standard output and is more difficult to get from a 2D flood simulation because the drainage also needs to be included in the schematisation (Wagenaar, 2012). During the 1993 Meuse flood, most drainage occurred because of the natural slope in terrain and therefore the 2D flood simulation implicitly includes most of the drainage because the discretised bed level is included. The flood duration can then be calculated by analysing the time-varying maps of the water depth and calculating for every cell the time between the moment a cell is inundated and the moment the cell is dry again. However, some cells in the digital elevation map in WAQUA are surrounded by cells that have a higher elevation. These cells do not drain in the 2D flood simulation and are still inundated at the end of the simulation. For these cells the flood duration has been calculated based on the change in water depth. If the water depth in a cell stays the same in the simulation for 24 subsequent hours the cell is considered dry at the moment this stable water depth is first reached.

Simulations were also ran with the same Meuse 1993 schematisation for design discharges with 1, 10, 50, 100, 250 and 1250 return periods. These discharges are based on HR2006 (Diermanse, 2004) and have discharges of respectively 1300, 2260, 2869, 3109, 3431 and 4000 m^3/s. The results of these simulations were combined to create a flood return period map for the Meuse 1993 situation. This map shows for each cell at what return period it first floods. This gives for each building an estimate of how often it has been flooded or threatened by floods in the past. This frequency of flooding for a specific house could be an indicator of how well prepared the inhabitants of a house are.

Figure 3.2 shows that large water depths occurred and that most of the area floods frequently. The majority of the houses is however located in the safest areas with the lowest water depths and highest return periods.

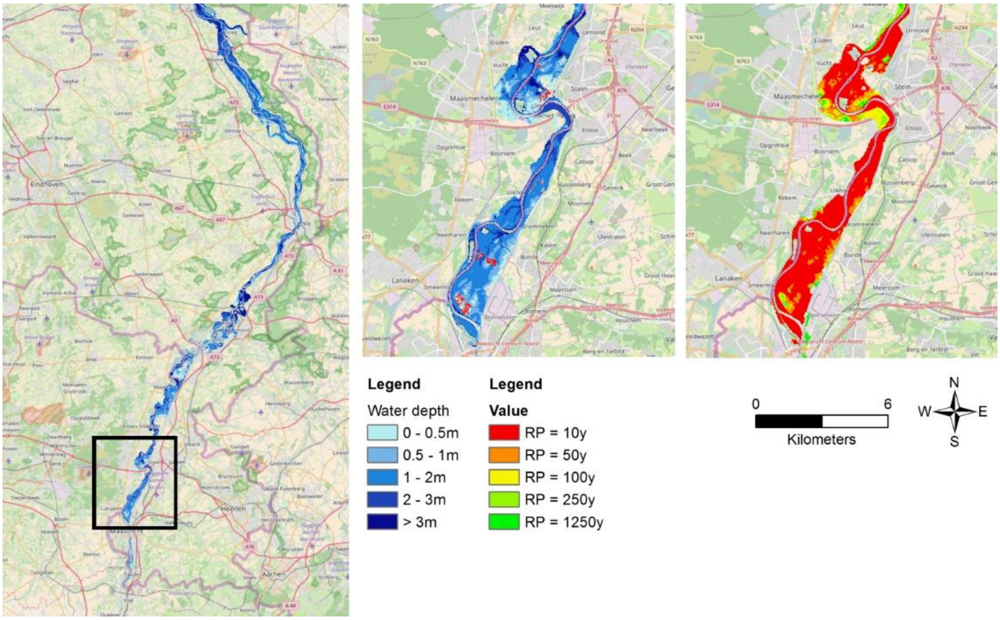


Figure 3.2: Left the simulated water depth for the entire study area in Limburg. In the center the simulated water depth and affected population (in red) for an example area. On the right the return period at which areas start flooding for the example area. The example area is defined in the box in the left picture. The scale bar corresponds to the example area.

These maps (water depth, flow velocity, flood duration and return periods) were linked to the original damage records using cadastre data. The data of the cadastre has exact building locations, postal codes, living area within the residential buildings, the building footprint area and the construction year. The building year was used to filter the data to find the building stock of 1993. Then, based on the building locations the 2D flood simulation results were linked to the cadastre data.

This combination of cadastre data and 2D flood simulation data is then used to make the link with the original flood damage records. First per postal code a list is made of the damage records in the postal code area and ranked based on the water depth in the original damage records. Then another list is made of the objects per postal code according to the cadastre and also ranked based on the simulated water depth. The cadastre objects combined with the 2D flood simulation data is then linked per postal code based on the water depth rank. This results in a join between the original damage records, cadastre data and 2D flood simulation results. Table 3.1 gives an overview of the available records in this combined dataset.

Table 3.1: Description of the variables in the flood damage dataset for the Meuse flood of 1993.

	Variable	*Unit*	*Source*	*Pearson correlation on total damage*
td	Total damage	Guilder (1993 value)	Original dataset[a]	1
sd	Structure damage	Guilder (1993 value)	Original dataset[a]	0.85
cd	Content damage	Guilder (1993 value)	Original dataset[a]	0.83
df	Water depth relative to floor	m	Original dataset[a]	0.18
dg	Water depth relative to DEM	m	Flood simulation[b]	0.18
bs	Basement	1=Yes, 2=No	Original dataset[a]	-0.04
dh	Detached house	1=Yes, 2=No	Original dataset[a]	0.08
hs	Household size	Number	Original dataset[a]	0.17
fv	Flow velocity	$m\ s^{-1}$	Flood simulation[b]	0.04
fd	Flood duration	h	Flood simulation[b]	0.05
rp	Return period	year	Flood simulation[b]	-0.09
ba	Building age	year	Cadastre[c]	0.01
la	Floor area for living	m^2	Cadastre[c]	0.04
fa	Footprint area building	m^2	Cadastre[c]	-0.02

[a] WL Delft, 1994

[b] 2D flood simulation data using WAQUA

[c] Basisregistraties Adressen en Gebouwen (BAG), version 2011 (Kadaster website).

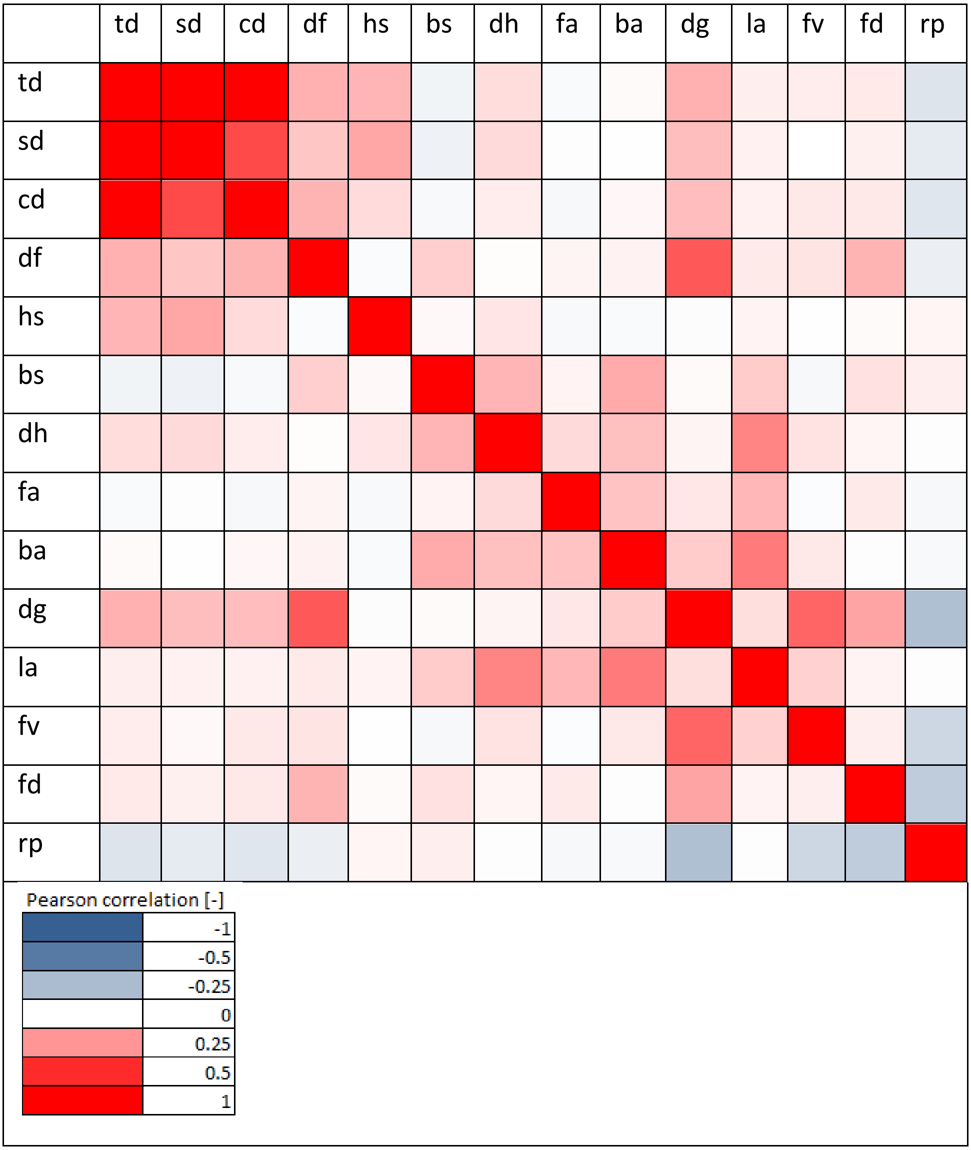


Figure 3.3: Correlation coefficients between the different variables. See Table 3.1 for a description of the abbreviations). Note: This table isn't applied in any of the supervised learning algorithms.

The method of joining cadastre objects with damage records within a postal code area based on water depth rank is error prone. The modelled water depth is on average 30 cm larger than the recorded water depth. This is possibly because the difference in reference level of both data sources as the recorded water depth is relative to the floor level and the modelled water depth is relative to the digital elevation map. Not all houses have the same floor elevation and both the recorded and the modelled water depth are uncertain, because of recording and model imprecisions. It is therefore likely that some damage records have been linked to the wrong object. However, errors will likely be limited, because the join on postal codes is accurate. Object and flood variables are generally similar for buildings within the same postal code area (e.g. houses within a street are typically similar to each other) so these errors are expected not to significantly disturb the general trends in the data. The errors are therefore considered acceptable given that the purpose of the dataset is only to build a flood damage model. If significant errors are present this would result in a reduced performance of the supervised learning algorithms on the test set. A relatively simple alternative to this water depth rank method is also applied. In this alternative, the average value at all building locations in the postal code area was assigned to each of the objects in the postal code.

3.2.2 Supervised learning algorithms

Several supervised learning techniques have been applied to the enriched dataset to build multi-variable flood damage models. The different supervised learning techniques all have different ways to generalize the training data in such a way that it can give useful predictions of the total damage.

These multi-variable flood damage models are be compared to two different reference models to assess the value of the enriched dataset and to assess the value of multi-variable flood damage models in general. Below the different supervised learning algorithms applied are described in further detail.

3.2.2.1 Regression: Root function

The first reference model only uses the square root of the water depth (see formula 1) to predict the flood damage. This model represents the damage functions commonly applied today in flood risk management studies because many damage functions have approximately the shape of a root function (e.g. Scawthorn, C., et al., 2006; Thieken et al., 2008; Penning-Rowsell et al., 2005; Sluijs et al., 2000). Merz et al. (2012) applied the same method to get a reference damage function. The purpose of this reference model is to see the benefits of using more data.

The root function (1) is fitted to the dataset in such a way that the coefficients c_1 and c_2 are optimised to get the smallest possible error based on the total damage (td) and water depth (wdf) data. The values of the coefficients are optimized for the best fit with the ordinary least squares method. This is done with the Python package SciPy.

$$td = c_1 + c_2\sqrt{wdf} \qquad (1)$$

3.2.2.2 Multi-variable linear regression

The second reference model uses multi-variable linear regression to fit a linear model to the data. This model represents more simple/traditional techniques to make a multi-variable model from data. The purpose of this reference model is to see the benefits of potentially better techniques to build multi-variable models from data. Multi-variable linear regression is for example used in Islam (1997) to make multi-variable flood damage models. Linear regression is used without transformations of the input variables, because there is no clear indication that in the data that there are non-linear relationships (for example see figure 3.1).

To ensure that the model captures general trends and doesn't fit too strongly to the observed data (overfitting) the LASSO technique is used. This technique determines the coefficients in such a way that a penalty is applied for increasing the coefficients and using the variables more. LASSO yields sparse models, so some coefficients will become zero, which means they are not useful for the prediction. Therefore, the LASSO technique is useful for variable selection. To make this work correctly the data is normalized before training the model.

The multi-variable linear regression was carried out with the Scikit learn library in Python (Pedregosa et al. 2011). LASSO requires an alpha parameter to be set which determines the height of the penalty applied. Several alpha values were tried (0, 0.5, 1and10). The model is very insensitive to the Alpha value (all formulations perform about equally well), an alpha value of zero performs best on all indicators. Therefore, it is not optimized further and the alpha is set to zero. When alpha is zero the method is equal to the ordinary least square method and no overfitting prevention is in place and LASSO is not necessary. This shows that overfitting is not an issue for relatively simple techniques such as linear regression with this dataset and number of variables.

3.2.2.3 Regression tree learning

Decision trees are a way to represent complex relationships between data and classes in a tree structure. A decision tree can be seen as a series of binary questions (nodes) leading to an answer in the form of a class (leaf). A question can be related to any variable at any value (e.g. is the water depth smaller than 0.5m).

A regression tree is similar to decision trees but instead of classes it results in real numbers. In theory, regression trees can be very large and have a separate leaf for each unique value in the dataset. However, it is more common to combine several similar unique values inside the same leaf and represent it with a summary statistic number (mean). In such a case the regression tree is an approximation of the relationship.

Regression tree learning algorithms can create optimal regression trees based on a dataset. In this paper the dataset consists of 4398 flood damage records (incomplete records are discarded) with 11 variables for each damage record (see table 3.1). The regression tree algorithm aims to split the dataset into subsets in such a way that the mean squared error (MSE) of the predicted total damage for all observations is minimized compared to the observed data. It does this by calculating the reduction for all candidate splitting variables according to their value and then picking the combination that maximises the MSE reduction ($\Delta\mathrm{I}$), this is shown in (2). n is the total number of observations in the node, $\mathrm{y_n}$ is the vector of observed target values in the node and $\bar{\mathrm{y}}$ is the mean of the target values in the node. $\mathrm{y_{nL}}$ and $\mathrm{y_{nR}}$ are vectors with the observed target values of the left and right group after the split and $\bar{\mathrm{y}}_\mathrm{L}$and $\bar{\mathrm{y}}_\mathrm{R}$ are the mean observed target value for the left and the right group. The regression tree is grown by repeating this process at each node of the tree. This has been done with the Scikit learn library in Python (Pedregosa et al. 2011).

$$\Delta I = \frac{1}{n}\left(\sum(y_n - \bar{y})^2 - \sum(y_{nL} - \bar{y}_L)^2 - \sum(y_{nR} - \bar{y}_R)^2\right) \qquad (2)$$

A regression tree algorithm keeps splitting the dataset into new branches until no more reductions in the MSE can be made. This can result in overfitting, which results in very large trees with only one data point per leaf. These very large trees are not a realistic representation of reality, and they typically perform badly when they have to predict the damage for a new data point that wasn't used for building the tree. There are several methods to prevent overfitting. The simplest methods require a minimum number of data points in a leaf or set a maximum number of nodes that the tree is allowed to contain. The disadvantage of these methods is that they sometimes don't build out a branch within the tree which at first doesn't look promising but which can make valuable homogeneity improvements deeper in the tree. A method called pruning is a more sophisticated method, in which the entire tree is first build with a subset of the data points, and then cut back based on its performance on data points that were not used for building the tree. The tree is cut back by removing the nodes by their performance improvement (least performing nodes first), the optimal pruning depth is than picked by testing the different pruning depths on the test set. This method was investigated in this research. This was done using *Matlab's 'Statistics and Machine Learning Toolbox'* (Matlab website), based on the work by Breiman et al. (1984), because the Python libraries do not support pruning. The MAE was applied as metric to find the optimal pruning depth. The performance of the pruning algorithm on this dataset was similar to a regression tree built with a combination of a minimum data point requirement per leaf and a maximum number of leaves (MAE with pruning in Matlab is 0.55 against 0.56 without pruning in Python). Therefore, the rest of the study was performed without pruning in the Scikit learn library in Python (Pedregosa *et al.2011*). Accordingly, the results shown do not include pruning.

3.2.2.3 Bagging regression trees

Another method to avoid overfitting and generally improve the accuracy of decision/regression trees is bootstrap aggregating, also called bagging. The idea behind the method is to resample the dataset multiple times and to build a new regression tree for each resampled dataset. This results in an ensemble of regression trees. The resulting flood damage is then the average of the ensemble of regression trees. Resampling is done by building several datasets by randomly picking records from the original dataset (each record is allowed to be used multiple times in the same dataset). Every resampled dataset therefore randomly leaves out a fraction of the observations and puts more weight on other observations because they are picked multiple times. Bagging regression trees also lead to probabilistic outcomes because the ensemble of trees can be seen as a probability distribution of the outcome.

3.2.2.4 Random Forest

A random forest is a more advanced variation of bagging regression trees. Apart from building multiple trees with resampled datasets it also randomly excludes a subset of variables at each decision split. This will result in an ensemble of regression trees each based on a different set of damage records and each leaving out a different number of variables at each decision split. For this paper the default settings of Scikit learn are applied, in our case this means 8 variables are left out at each decision split.

3.2.2.4 Bayesian Network

A Bayesian Network is a type of Probabilistic Graphical Model that represents a set of random variables and their conditional dependencies in a directed acyclic graph (DAG) structure. Each variable in the network may be observed or represented as a prior probability distribution and dependencies between variables are represented with edges representing joint probability distributions. The edges in a Bayesian Network are directed which means there is a direction in which the influence of one variable flows to the other. From this network, inference can be done in order to use knowledge of one variable to make predictions about other variables.

Bayesian Networks and Probabilistic Graphical Models in general are used in many different fields, such as bioinformatics (e.g. Mourad et al. (2011), image processing (e.g. Sudderth & Freeman, 2008) and speech recognition (e.g. Bilmes, 2002). Recently, they have also been applied to flood damage modelling (Vogel et al., 2014; Schröter et al. 2014; Van Verseveld, 2014). Schröter et al. (2014) found that their performance is often better than that of the different types of tree methods. Furthermore, a Bayesian Network can give its result as a probability distribution and does not require information about each variable in order make predictions. If fewer variables are available, the Bayesian Network handles this by adjusting the probability distribution of the outcome. This makes it ideal for transfer of models to other locations where less data is available than for the location where the model was originally based on. Furthermore, it returns (for each object) probability distributions rather than deterministic values, which is valuable for assessing uncertainties within the damage model estimates.

A Bayesian Network can be discrete, continuous or a combination. In this paper fully discrete Bayesian Networks are used, in which all variables are discretized into bins. Given a network the probability of particular set of discrete variable values can be calculated with the following formula:

$$P(X_i, \dots, X_n) = \prod_{i=1}^{n} P(X_i | parents(X_i)) \quad (3)$$

Where X_i are the variables and $parents(X_i)$ is the set of variables directed to X_i. The probability of a single variable value can be obtained by taking the sum of all the probabilities that contain the variable value of interest. The conditional probabilities are stored in conditional probability tables (CPTs). These tables show, for each combination of parent variable values, the probability of each possible output value.

A data-driven Bayesian Network can derive all its CPTs from the data and even derive its graph structure from the data. For this paper, two Bayesian Networks were made: A data-driven Bayesian Network with both the graph structure and the CPTs derived from the dataset and an expert network where the graph structure was estimated in an expert session but the CPTs were derived from the dataset. All calculations were done with a Python library called libpgm (Cabot, 2012). This library follows the methodology described in Koller and Friedman (2009).

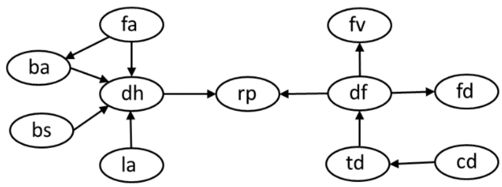
The CPTs are learned with maximum likelihood estimation. This method estimates the (joint) probability distributions based on the number of observations. The discretisation assumptions have an impact on the maximum likelihood estimation. If the variables are discretised into a large number of bins more possible combinations of states are possible. These combinations of states grow exponentially with the number of bins of the parent variables. A too fine discretisation therefore quickly leads to more possible states than available data points. This results in a poor performance of the maximum likelihood estimation. Koller and Friedman (2009) call this one of the key limiting factors in learning Bayesian Networks from data. A too coarse discretisation on the other hand is also not desirable because it limits the precision of the Bayesian Network. For this study a balance was found by trying several discretisation resolutions until the best result was found based on the MAE criterion.

Discretisation was done by splitting the data into bins with an equal number of data points in each bin. This works better than making equal sized bins because of the large extremes in especially the damage data. Equal sized bins would either increase the number of bins, which is detrimental to the maximum likelihood estimation (having bins that contain no observations), or the bins would be so large that a majority of the data points would end up in the same bin, which would limit the Bayesian Network performance. The number of bins per variable was chosen based on the performance of a test set on the MAE criterion. This was done by varying the discretisation of the most important variables until the smallest error was found. For the Bayesian Network with the data-driven structure the number of bins chosen was slightly larger, because the network is less complex than the expert network.

The performance of the Bayesian Network on the testing data can be sensitive for discretisation. There are two possible alternatives for the discretisation method applied in this paper: An optimisation algorithm could be applied to determine the optimal discretisation, or a continuous Bayesian Network could be used (Friedman and Goldszmidt, 1996). Apart from solving the discretization problem the advantage of a continuous Bayesian Network is that it would probably perform better in predicting extreme values but a disadvantage is that the Bayesian Network is restricted to specific families of parametric probability distributions (Friedman and Goldszmidt, 1996). An optimization algorithm for the discretization can minimize the error produced by the discretizing but does not solve the fundamental problem of having too few data points.

The data-driven structure is also learned with the libpgm Python library. This library is using a constrained-based approach for structure learning, as is described in Koller and Friedman (2009). In a constrained based approach the structure is learned by calculating dependencies and conditional dependencies among the variables. When two variables are dependent regardless of what they are conditioned by, an edge (connection) is formed. The algorithm follows this procedure to create the entire network. The result is shown in figure 3.4 (left).

As an alternative to the data-driven structure a structure was also made in an expert meeting involving several Deltares flood damage/Bayesian Network experts (see acknowledgements). In the expert meeting the network was constructed based on a combination of expert judgement/logic and with the knowledge of figure 3.4 in this paper. The experts focused mainly on edges that they thought are relevant for estimating the flood damage. The result is shown in Figure 3.4 (right).



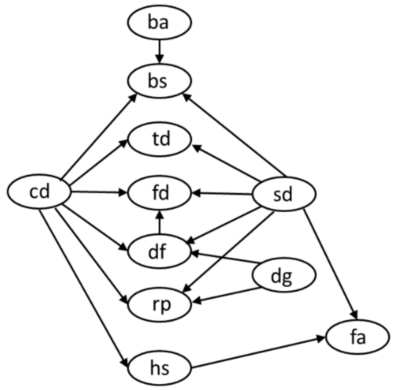


Figure 3.4: Bayesian Network learned from data (left) and Bayesian Network constructed by experts (right). Note that not all variables are used in the network and that the total damage in the expert network was not used in the actual calculations (because its simply the sum of structure and content damage)

The relationship between the total, structural and content damage is known and not probabilistic: total damage = structure damage + content damage. Also, in our case the structure damage, content damage and total damage are always all dependent variables. Therefore, using a Bayesian Network to model this exact definitional relationship could only introduce extra errors and not add anything extra explanation. Therefore in the expert network it has been decided not to use the total damage variable. Instead the total damage is calculated as the sum of the expected value of the structure and the content damage. In the data-driven network the structure damage was not included by the algorithm. Therefore, the total damage variable itself is used for the data-driven network.

The advantage of an expert based network is that experts focus on the connections that matter most rather than on all possible connections. Furthermore, experts can include connections that are not found in this dataset but are expected to exist in theory or in an independent test set. The advantage of a learned network is that new and previously unknown relationships between variables can be discovered.It is expected that the Bayesian Networks in this manuscript are not very sensitive to overfitting during the CPT learning. Koller (2008) only mentions overfitting in the maximum likelihood estimation of Bayesian Networks in relation to discretization that is too fine and offers no techniques to counter overfitting in the maximum likelihood estimation. This expectation that overfitting isn't an issue was tested by testing the Bayesian Network on its own training data. If overfitting is an issue the model should do much better in predicting its own data then in predicting new data. This isn't the case (for the expert model) the MAE is even slightly worse when calculated on its own data (0.622), the correlation coefficient and R^2 are only slightly better (0.24 and 0.04) and only the mean bias error (MBE) is significantly better (-0.015). See results section for comparison.

3.2.3 Variable importance

In order to investigate the value of more data it is interesting to study the contribution of the different variables to the prediction accuracy. This can be done with bagging trees and the random forests methods. This importance can be calculated as the (normalized) total reduction of the mean square error brought by the different variables as achieved during the training of the models. This can be used to compare the relative importance of the variables among each other. This feature importance can be calculated for all the regression trees in the ensemble and a general importance is computed by the sci-kit learn library by taking the average of the feature importances in the tree. This was applied in this study for the bagging trees. The variable importance has been separated for predicting the importance of the total damage, structural damage and the content damage. For the calculation of the variable importance the dataset is used in which the average per postal code is used for the new variables. The water depth rank is not used because it could transfer some of the importance of the original water depth value to the new variables.

Another way to study variable importance is with the LASSO technique in multi-variable linear regression. LASSO can drop unimportant variable coefficients to zero. If a variable is dropped to zero it means the variable is less important.

3.3 Results

3.3.1 Model comparison

The different models are tested on a test set that was not used for training the models. Four indicators are used to rate the performance of the models: Mean Absolute Error (MAE), Mean Bias Error (MBE), the Pearson correlation coefficient, and the coefficient of determination (R^2). The MAE is the mean absolute error divided by the average damage (see table 3.2), so a smaller MAE is a better model. The MBE is the average error, this differs from the MAE in that an overestimation is able to correct for an underestimation and the other way around. A low MBE shows that the sum of a large number of predictions will probably be very accurate. The Pearson correlation coefficient is a measure of the linear dependence between two variables. This measure is used to compare the predicted damages with the actual damages in the test set. A Pearson correlation of one means a perfect correlation, zero means no correlation and minus one a perfect inverse correlation. R^2 is the predictive capacity of a model compared to just using the average damage as a prediction. If the R^2 is zero it means the independent variables add no predictive capacity compared to just using the average. When R^2 is 1 it means the independent variables can explain all variation in the dependent variable. Table 3.2 shows the results for the different models.

Table 3.2: The formulas for the error metrics. Note, in later chapters these metrics aren't divided by the mean damage. $D_{predicted}$ = predicted damage for a house, D_{actual} = damage that actually occurred in the house. $\mu_{damage,actual}$ the mean actual house damage.

Error metric	*Formula*
Mean Absolute Error (MAE)	$MAE = \frac{\sum \lvert D_{predicted} - D_{actual} \rvert}{\mu_{damage,actual}}$
Mean Bias Error (MBE)	$MBE = \frac{\sum (D_{predicted} - D_{actual})}{\mu_{damage,actual}}$

Table 3.3: Results of different models for four indicators: MAE, MBE, R^2 and correlation coefficient. The models had access to all variables (except for the root function). The version of the dataset with the water depth rank join between the old and the new variables is used .

Calculation	*MAE*	*MBE*	*R^2*	*Correlation coefficient*
Root function	0.612	0.194	0	0.15
Multi-variable linear regression	0.578	0.055	0.07	0.27
Regression tree	0.561	0.065	0.03	0.31
Bagging regression tree	0.504	0.061	0.15	0.38
Random forest	0.508	0.054	0.16	0.39
Data-driven Bayesian Network	0.629	0.525	0	0.21
Expert Bayesian Network	0.607	-0.08	0.03	0.21

Table 3.3 shows that given that the models can use all data, random forest and bagging regression trees perform best and equally well. These two methods reduce the MAE by 12% compared to a reference model using the same data (multi-variable linear regression). Bagging regression trees and Random Forest do perform significantly better than normal regression trees, as was also noted by Merz et al. (2013) for flood damages in Germany. Random Forest and Bagging regression trees also outperform the Bayesian Networks. The normal regression tree also works better than the Bayesian Networks. This contradicts earlier findings by Schröter et al. (2014), who found that in most cases Bayesian Networks outperformed the regression trees. Schröter et al. (2014) did however have a very different dataset from the one applied in this study.

Many explanations are possible for the relatively poor performance of the Bayesian Networks. The discretization of the data is a possible problem. Some trends could be too subtle to be captured by the rough discretization, but not enough data points are available for a more precise discretization. Perhaps there still is some space for improving the discretization, for example by applying an optimization algorithm to pick bin definitions in such a way that the available information is applied optimally (Vogel et al. 2012 applied such an algorithm). Another possible reason is that Bayesian Networks might be more sensitive to low quality data in combination with a small dataset. Some of the CPTs applied in the Bayesian Networks here are large and conditional probabilities are based on a relatively small number of observations. Some wrong observations may then have a relatively large impact on the damage prediction.

In the data-driven network the variable of interest (total damage) in our test is only influenced by the water depth. This is because the water depth relative to the ground floor is known while the content damage is not known, so the known water depth blocks all the influence of other variables and the unknown content damage has no influence because it is unknown (it is a target variable). The data-driven Bayesian Network is therefore in our test in practice only dependent on the water depth. So the structure learning decides to ignore the other variables when the water depth relative to the ground floor is available. This is probably because the data-driven structure algorithms finds all variables equally important and therefore draws only the most important edges (connections) regarding the total damage. Other methods (e.g. as described by Riggelsen, 2008) for structure learning might be able to give better results.

The multi-variable linear regression reference model does a good job on the MBE but is clearly weaker on the other performance indicators, which shows that for predicting aggregate damages for e.g. policy studies, the more complex methods are less beneficial. This is different in cases where individual building damages are important, for instance for insurance rating purposes. The reference root function has a very large bias compared to the other models. This is probably because the shape of the root function is inappropriate for this flood event.

3.3.2 Benefits of more data

The models were trained with different numbers of variables to see whether the additional data is valuable. As expected, the best performing model with a high number of variables always performs significantly better than the best performing model with fewer variables (see table 3.4). More data therefore seems to add potential value to the damage prediction despite the possible quality issues in the additional data. The MAE of the best performing model with only the water depth (regression tree) can be reduced by a further 14% by the best model using all data (Random Forest). The MAE of the root function fitted to the data (representing common practice) can be reduced by about 20% using the Random Forest with all data.

The method to join the extra data with the original data based on water depth rank is not effective. Just taking the average value per postal code appears to work better.

The water depth rank probably sometimes assigns extreme variable values to the wrong objects which disturb some correlations in the data.

Table 3.4: The best performing model based on the MAE indicator with different number of variables.

Variables	*Method*	*MAE*	*MBE*	R^2	*Correlation coefficient*
Only water depth	Regression tree	0.564	0.071	0.08	0.306
Only original variables (waterdepth, household size, detached house, basement)	Bagging trees	0.551	0.052	0.07	0.345
All variables (water depth rank join)	Random Forest	0.508	0.054	0.16	0.394
All variables (average postal code join)	Random Forest	0.488	0.035	0.17	0.41

3.3.3 Variable importance

The total importance of variables that were added in this study is about 30% (figure 3.5), that means that 30% of the error reduction during the training of bagging tree model originates from variables that were added to the dataset. The added variables therefore clearly help to improve the prediction accuracy. This assessment was done without the water depth ranking join because this could assign some of the importance of the original water depth to the modelled water depth. The original water depth is by far the most important variable. Construction year is an important variable for the structure damage but not for the content damage. This is as expected. Household size is quite important for the structural damage but insignificant for the content damage. This is less obvious but it could be that large families live on average in larger houses but do not have much more valuable contents on the ground floor. Return period is an important variable for both the structure and the content damage. This was also expected because the population in areas that flood more frequently are expected to have more flood experience, thus resulting in better preparedness and lower damages. This effect is visible in the data, with return period having an importance of about 10%.

For the best fitting multi-variable linear regression model (LASSO alpha=0) no variables are dropped. Only when the alpha is increased to 10, 5 variables are dropped, however this also causes a slight drop in model performance (MAE goes from 0.578 to 0.588). The dropped variables are: Building footprint, building age, living area, flood duration and flow velocity. From these dropped variables, two have a significant importance in the bagging tree variable importance assessment. These are building age and living area. It could be that those variables are more important in non-linear models.

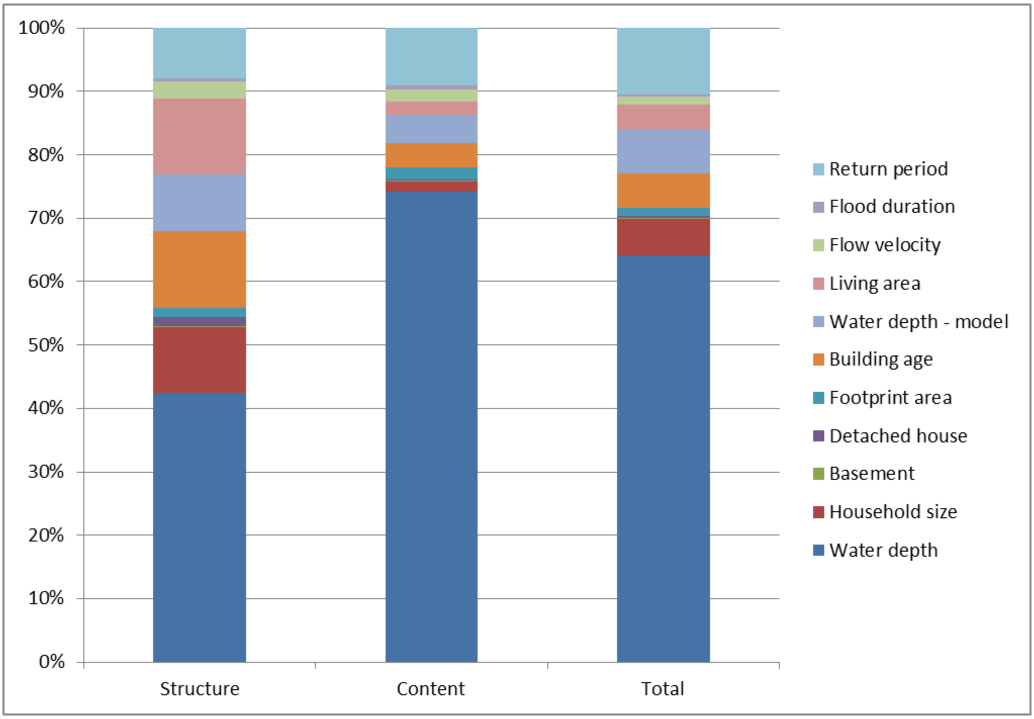


Figure 3.5: Variable importance: The contribution of different variables in reducing the error in the bagging regression trees (the chart follows the order of the legend).

3.4 Discussion and conclusion

Additional data improves flood damage modelling relative to a test set, even if this data comes from a collection of different sources and is of limited quality (error prone). The supervised learning algorithm is also important. Given the same data there are large differences between the algorithms. Random Forests and bagging regression trees perform significantly better than normal regression trees and multi-variable linear regression. The Bayesian Networks perform poorly compared to any of the tree based methods.

Our current approach doesn't show that the additional variables are beneficial for the Bayesian Networks. However, because the tree methods can benefit from the additional data it is likely that in some cases Bayesian Networks could also. The poor performance of the Bayesian Networks contradicts earlier studies (Schröter et al., 2014) and could be due to the discretization method, quality of the expert network, network learning algorithm or problems with data quantity or quality.

The test set that was applied in this paper for the validation of the model, was randomly selected from the data and consistently applied among all models. The accuracy of the indicators for model performance could perhaps be further improved through some form of cross-validation. Also the tweaking of different models could become more accurate if cross-validation was used instead of validation on a single test set only. For example, the optimization of the stop criteria for tree based models and the alpha value in the LASSO method for the multi-variable linear regression could be improved that way. Expectations are that this would cause minor improvements in results but that it would not influence the conclusions of this paper.

This paper did not address another benefit of Bayesian Networks, Random Forest and Bagging trees, which is the incorporation of uncertainty. Bayesian Networks do this explicitly in the method and for Bagging Trees or Random Forest each tree can be seen as a possible damage estimate and together the trees represent a probability distribution.

The methods applied in this manuscript provide an uncertainty estimate for a single object. For policy decision making it is often useful to aggregate these uncertainty estimates to a total uncertainty for the entire flood event. This can be done with the assumption that all objects are perfectly correlated to each other (one tree will apply to the entire event but what tree is uncertain), or with the assumption that all objects are independent of each other (each object will have a different tree but what tree is uncertain). Both assumptions are however not completely correct (Wagenaar et al., 2016). The Bayesian Network framework might offer a middle way to model this correctly. If each object has a copy of the original Bayesian Network, and these Bayesian Networks are linked together based on the location of the objects, it can be explicitly taken into account that nearby objects are more likely to have similar damages. This could be an argument to prefer Bayesian Networks over tree based methods in the future.

The dataset applied in this paper had many limitations. The most important limitation is that the exact locations of the objects are unknown. Because of this, it was difficult to link building and flood characteristics to damage records. An attempt to do this by using water depth rank performed worse than just using the average variable values per postal code. Despite this limitation, the added data still produced significantly better damage estimates. Another problem with the dataset is the unknown manual adjustment to an unknown share of data (rental residential buildings) for the structural damage records. These actions may have introduced a relationship between structural damage and some of the originally recorded variables that wasn't there in reality. This could in theory cause a slight overestimation in the prediction performance of the models on the test set. This effect on the results is however expected to be small, because most of the prediction improvements came from adding variables that were not available for doing the manual actions in 1994.

This study applied absolute damages rather than relative damages. This requires the supervised learning algorithms to implicitly also predict information about the values at risk besides the vulnerability. The algorithms can do this with variables such as living area, footprint area, building year and household size. This seems less error prone and better than estimating such values at risk with general rules of thumb based on assumptions about construction costs and content value. Such assumptions could cause extra errors, and therefore in this study absolute damages were used.

This paper trained flood damage models on just a single flood event. Ideally training data should consist of multiple events so that the spectrum of possible damages which the model is trained upon is larger. Especially for the transfer to other areas this would be important. Models that are trained on a single event could overfit on this event and this problem would not show up if the model is tested with data from that same event (even if this specific data wasn't used for training the model). A good example of this appears in the good performance of the regression tree based on only the water depth versus the fitted root function based on only the water depth. The root shape of a damage function which many expert models use (see section 2.2.1) and which makes physically sense, is performing much worse than a more flexible model that can adjust to other relationships between damage and water depth. This is explained by figure 3.6 which shows a downward sloping damage function after 90cm of water depth, a shape very different from damage functions normally found in the literature. The root function model therefore starts producing large errors after 90 cm while the regression tree can capture this trend well. This downward sloping makes physically no sense but could be explained by other variables such as return period. Return period could be a proxy for flood experience and better preparation because houses that experienced large flood depths in 1993 are probably on lower ground and also experience floods in general more often. In this case return period could therefore dampen the effect that greater water depths cause more flood damage. For some water depth ranges the return period could even cause this downward sloping damage function (see figure 3.6). This relationship is probably not true for other types of events, for example large flood depths due to dike breaches. So in that sense, the regression tree is overfitting on this single flood event.

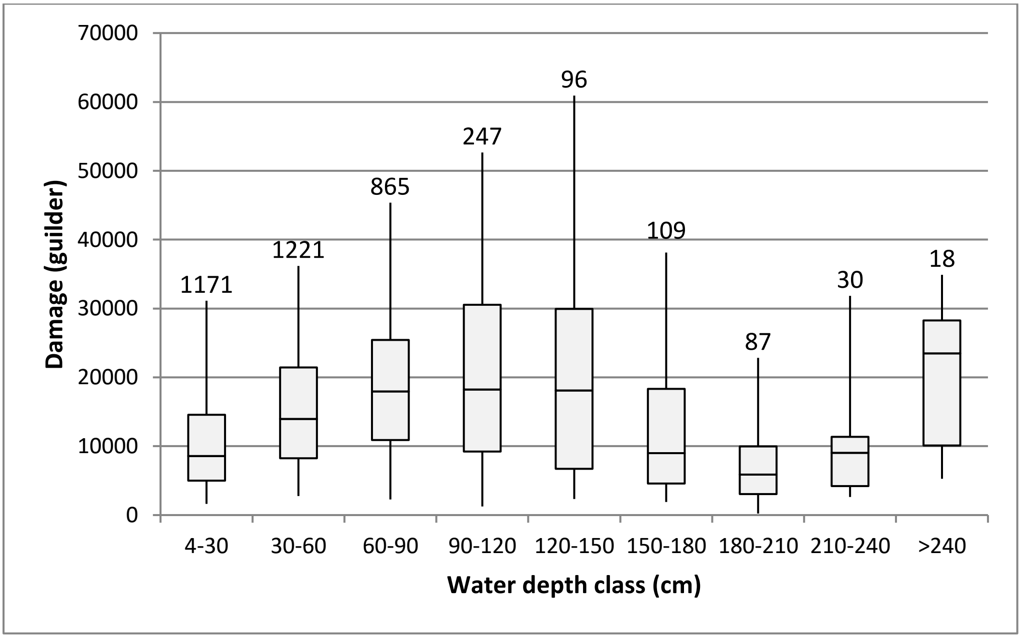


Figure 3.6: Box-plots of the Meuse flood of 1993 per water depth class. The box shows the 25-75% interval and the lines show the 5-95% interval. The line in the middle of the box shows the median value. The labels on top of the plots show the number of observations per water depth class.

Supervised learning can help to create and improve flood damage models. They have many theoretical advantages over deterministic damage functions based on only the water depth. The application of supervised learning in flood damage modelling remains challanging in practice, because of limited data availability. In this paper we utilized different data sources compared to previous studies to acquire this data and showed that also on this dataset the methods are beneficial, especially the tree based methods. Future work may merge available datasets from different events and from different countries in order to develop a model that can be applied using several hazard variables, and which also works in circumstances outside areas for which flood damage data is available.

Acknowledgments

We thank our colleagues Kathryn Roscoe for advice on the Bayesian Networks and our colleagues Karin de Bruijn and Marcel van der Doef for their input in constructing the expert Bayesian Network. We would also like to thank the editor (Dr. Thaler) and the reviewers (one anonymous reviewer and Dr. Fuchs), their detailed constructive comments and suggestions helped to substantially improve this paper. This research has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement number 641811 (Improving predictions and management of hydrological extremes – IMPREX), see also http://www.imprex.eu.

4. TRANSFERABILITY OF MULTI-VARIABLE FLOOD DAMAGE MODELS

This chapter is based on the paper "Regional and temporal transferability of multi-variable flood damage models" written together with Stefan Lüdtke(GFZ), Kai Schröter (GFZ), Laurens Bouwer (Climate Service Center Germany) and Heidi Kreibich (GFZ). It is published in the journal Water Resources Research (WRR). The reference is: Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L.M., Kreibich, H., 2018. Regional and Temporal Transferability of Multivariable Flood Damage Models. Water Resources Research. Volume 54, Issue 5, Pages 3688-3703. https://doi.org/10.1029/2017WR022233

Abstract

Reliable flood damage assessment is important for decision-making in flood risk management. Flood damage assessment is often done with damage curves based only on water depth. These depth-damage curves are usually developed based on data from a specific location and specific flood conditions. Such depth-damage curves tend to be applied outside the scope of their validity. Validation studies show that in such cases depth-damage curve are not very reliable, probably due to excluded influencing variables. The expectation is that the inclusion of more variables in a damage function will improve its transferability. We compare multi-variable models based on Bayesian Networks and Random Forests developed on the basis of flood damage datasets from Germany and The Netherlands. The performance of the models is tested on a validation sub-set of both countries' data. The models are also updated with data from the other country and then tested again. The results show that the German models (BN/RF-FLEMOps) perform better in the Netherlands than the Dutch models (BN/RF-Meuse) perform in Germany. This is probably because the FLEMOps models are based on more heterogeneous data than the Meuse models. The FLEMOps models, therefore, are better able to capture damages processes from other events and in other locations. Model performance improves via updating the models with data from the location to which the model is transferred to. The results show that there is high potential to develop improved damage models, by training multi-variable models with heterogeneous data, for example from multiple flood events and locations.

4.1 Introduction

Flood risk management is becoming increasingly risk-based, and flood damage estimation is therefore increasingly important in flood risk assessment studies (Merz et al., 2010). Reliable flood damage estimates support policy makers in making sound cost-benefit analyses when assessing and prioritizing risk reduction measures (Kreibich et al. 2014). An example in which flood damage estimates played an essential role is the Dutch Delta Programme, where they were used to determine the optimal protection level for the levee systems in the Netherlands (Kind, 2013; Van der Most, 2014).

Flood damage models typically estimate damage based on water depth and average building value for different types of buildings. When validated, such flood-damage models often perform poorly (e.g. Jongman et al., 2012) and there exists a lot of unexplained variation among damage functions in the literature (Wagenaar et al., 2016; Gerl et al., 2016). Merz et al. (2004) showed that water depth alone explains only a part of the variation among flood damage observations. Therefore, the most likely reason for the differences between damage functions is that these models contain implicit assumptions about variables not included (Wagenaar et al., 2016). Examples of such variables are: Flood duration, flow velocity, precautionary measures on household level, contamination of the flood water and household size. Including other additional variables in flood damage models can substantially improve the reliability of flood damage estimates (Wagenaar et al. 2017; Kreibich et al. 2017; Schröter et al. 2014). Thieken et al. (2008) and Kreibich et al. (2010) found that contamination and precautionary measures are important for predicting the flood damage in Germany, and introduced rule based loss estimation models (FLEMOps, FLEMOcs) with correction factors to adjust for these variables. Applications and validations of these models on the micro- and meso-scale are described by Seifert et al. (2010) and Falter et al. (2015, 2016). Merz et al. (2013) introduced tree based models to learn multi-variable damage models from data. Vogel et al. (2014) introduced the application of Bayesian Networks for the same purpose. These multi-variable flood damage models perform better than traditional flood damage models, particularly in a spatial and temporal transfer setting (Schröter et al., 2014).

Detailed data on flood damages are only rarely recorded, for instance after some flood events via surveys in Germany, France and Italy (Kienzler et al., 2015, Poussin et al., 2014, Molinari et al., 2014). Thus, for low-probability events, data often is not available or is outdated. In the application of flood damage models, a spatial and/or temporal transfer of models is therefore often necessary (Cammerer et al. 2013). For example, the Dutch standard model for flood damage estimation SSM2015 is based on transferred damage functions in both space (other countries) and time (1953 data) (De Bruijn et al., 2014). Other examples of such studies that transfer damage models are: Dahm et al. (2017), Tollenaar et al. (2016), Bouwer et al. (2017). This however can easily lead to errors, as the depth-damage functions are often not valid for the area to which they are transferred.(Papathoma-Köhle et al., 2011; Meyer et al., 2013) For transferability of flood damage models, it is therefore advantageous to describe the complexity of the damage process with multiple variables. A larger number of variables in the model could account for some of the implicit assumptions in the flood damage model and therefore make the model better transferable to other areas. Schröter et al. (2014) found that a complex multi-variable flood damage model outperforms simple models, when both are transferred to other areas. However, such comparisons are rare, and so far no study has compared the transfer of multi-variable damage models between two countries with different types of datasets.

The aim of this study is to test how well multi-variable models perform in a temporal and spatial transfer, using data originating from different countries with different collection methods. We used (1) A German dataset based on telephone interviews conducted after several different flood events in Germany (Kienzler et al., 2015; Thieken et al. 2016) and (2) a dataset from The Netherlands based on compensation data from the 1993 Meuse flood in the Netherlands (Wind et al., 1999). Both datasets are used to train damage estimation models based on Random Forests and Bayesian Networks which are then verified using the other dataset. This study will focus entirely on residential damages relative to the building value.

4.2 Data and Methods

4.2.1 Data sets

4.2.1.1 German flood damage data

The German data set contains flood damage data collected through surveys after floods in 2002, 2005, 2006, 2010, 2011 and 2013. Each flood event is briefly described at the end of this section. Four surveys were carried out, using computer-aided telephone interviews, with private households which had been affected by the 2002, 2005/2006, 2010/2011 and 2013 floods in Germany. Computer-aided telephone interviews were undertaken by polling institutes with the VOXCO software package (www.voxco.com). The person with the best knowledge of the flood damage was interviewed. The surveys contained about 180 questions addressing a broad range of topics: flood impact (e.g., water depth in highest affected floor, perceived flow velocity, flood duration at building), flood warning, emergency measures, evacuation, cleaning up, building characteristics (e.g. building type and age, floor area for living, footprint area of building, availability of basement), damage to household contents and building, recovery, precautionary measures (e.g. which measures were implemented before the flood), flood experience (e.g. how many floods experienced before event and how long ago, experience with flood damage), and socio-economic variables (e.g. household size). To avoid errors as much as possible, only meaningful answers were accepted by the system. Wherever possible, answers were cross-checked. In case of conflicting answers, the interviewee was informed about a contradiction and prompted to clarify the situation. On the basis of the information provided in the survey, indicators for flow velocity (score from 0=still to 3=high velocity), precautionary measures (score 0=no measures undertaken to 38=many, efficient measures undertaken) and flood experience (score 0=no experience to 9=recent flood experience) were developed as described in Thieken et al. (2005). On the basis of information provided about water depth and building characteristics, the water depth relative to the ground level was calculated, with negative values in case of basement flooding only. The relative building and contents damage were calculated by dividing the actual damage as given in the survey by the building and contents values (replacement costs as at the event), which were estimated by using valuation methods of the insurance industry (Dietz, 1999).

Other data was also used or collected in addition to the survey. Information from affected communities, flood reports, press releases, as well as with the help of flood masks derived from satellite data (DLR, Centre for Satellite Based Crisis information, www.zki.caf.dlr.de) were used to compile lists of affected streets. These provided the basis for generating property-specific random samples. Return periods of the flood at the affected residential buildings were estimated on basis of the annual maximum series of discharge of all gauges in the study areas using extreme value statistics as described by Elmer et al. (2010).

The variables used for this study are provided in Table 4.1. Further details about the surveys and the data processing are published by Thieken et al. (2005; 2017) and Merz et al. (2013). In total, data of 4,368 households are contained in the dataset, with the following distribution across events: August 2002: 1,697 households; August 2005: 305 households; April 2006: 156 households; August 2010: 349

households; January 2011: 209 households; June 2013: 1,652 households. Cases with missing values were excluded for the analyses, resulting in 1456 complete records for building damage and 1324 for contents damage.

In August 2002, a cyclonic depression from the southern direction led to an extreme flood event in Germany, Austria, the Czech Republic and Slovakia. Record-breaking precipitation occurred; for instance 312 mm within 24 hours at the gauge Zinnwald-Georgenfeld in the Ore Mountains, Germany (Ulbrich et al., 2003; Engel, 2004). Highly dynamic floods occurred in the Ore Mountains, e.g. at the rivers Mulde, Weißeritz, Schwarze Elster. Discharge return periods were estimated to be about 150-200 years at the Dresden gauge on the river Elbe, 200-300 years at the Mulde River, and 100-300 years at the Regen River, a left-bank tributary of the Danube (Ulbrich et al. 2003; IKSE, 2004). The flood caused 21 fatalities and overall financial damages of 11.6 billion Euros (price level 2002) in Germany (Thieken et al., 2006).

In August 2005, a considerable flood affected the German part of the Danube catchment. Cyclone 'Norbert' induced prolonged rainfall with notably high amounts within 12 to 24 hours in Switzerland, northern Italy, Austria and southern Germany, such as 216 mm in 24 hours in Balderschwang, Germany (LfU, 2007). The alpine foothills were affected by flash floods. Inundations occurred both along the river Danube and its southern tributaries. Return periods were classified to less than 100 years at the Iller, Schmutter, Amper, Inn and Isar rivers and to 20 to 50 years at the rivers Lech, Loisach and Mangfall. At the Danube River, the highest return periods occurred at the cities of Ingolstadt and Kelheim in the range of 20 to 50 years (LfU, 2007). The total economic damage was estimated to be about 175 million Euros (price level 2005) in Germany (Kron, 2009).

The flood in the spring of 2006 occurred mainly in the Elbe catchment. High amounts of water were stored as snow at the beginning of 2006 in the upper Elbe catchment (Korndörfer et al., 2006). At the end of March, heavy rainfall occurred and temperatures rose rapidly from 5 to 15 °C leading to a complete snowmelt also in the upper parts of the middle hills (BfG, 2006). At the Dresden gauge the return period was estimated to 15 years (Kreibich and Thieken, 2009). But the flood situation downstream of the Havel confluence was comparable to or even worse than in August 2002. For instance, at the Neu Darchau gauge, the flood discharge of 3,600 m3 s-1 in 2006 was the second highest in 100 years and exceeded the 2002 flood discharge of 3,400 m3 s-1 (BfG, 2006). The total resulting damage in Germany was estimated at 120 million Euros (price level 2006) (Kron and Ellenrieder, 2008).

Three heavy rainfall events in August and September 2010 resulted in extreme floods in the Odra and Elbe catchments (Walther et al., 2013). The flood situation was aggravated significantly due to the breach of the dam Niedow at the Witka River, which is a tributary of the river Lausitzer Neiße, on 7 August (Jelonek et al., 2010). In the upper parts of the Schwarze Elster and Spree catchments, the highest peak flows occurred at the beginning of August with return periods of up to 500 years at the Spree and up to 200 years at the Schwarze Elster. At their lower reaches, the highest flows occurred at the end of September with return periods of 50 to 100 years (Walther et al., 2013). Particularly high damage occurred in the upper reaches of the Lausitzer Neiße and Spree as well as at the Mandau River. The total resulting damage in Germany was reported to be 839 million Euros (price level 2010) (EC, 2014).

Processes leading to flooding in January 2011 were comparable with the flood in 2006. Due to massive snowfall in winter a lot of water was stored as snow. A temperature increase and heavy rainfall led to snow melt and a first increase in river discharges between 5 and 6 January 2011. In the following days, between 12 and 14 January 2011, large-scale, intense rainfall fell on already saturated soils which led to a second flood wave with water levels above the flood warning levels at many gauges (Axer et al., 2012). Nearly all large catchments in Germany were affected, e.g. the catchments of the Rhine, the Danube, the Weser and the Elbe (Axer et al., 2012). Particularly high discharges occurred at the rivers Main and Saale and in the upstream part of the Weser catchment. The total damage was estimated to be more than 100 million Euros in Germany (price level 2011) (Axer et al., 2012).

The flood in June 2013 was mainly driven by the combination of high catchment wetness due to a strong rainfall anomaly during the month of May and spatially extended high but not extraordinary precipitation (Merz et al., 2014; Schröter et al., 2015). Nearly all main river basins in Germany were affected, but particularly severe flooding occurred along the Danube river in the federal state of Bavaria and along the Elbe River and its tributaries Saale and Mulde in the federal states of Saxony and Saxony-Anhalt (Schröter et al. 2015). In Passau, the highest water level since 1501 was observed, due to the superposition of the flood waves from the Inn and Danube rivers (Blöschl et al. 2013; DKKV 2015). Due to the large spatial extent of flood peaks with high magnitudes, the June 2013 flood was in hydrological terms the most severe flood in Germany at least for the last six decades (Schröter et al. 2015). The flood caused 14 fatalities and overall financial damages of about 8 billion Euros (price level 2013) (Thieken et al. 2016).

4.2.1.2 Dutch flood damage data

The Dutch dataset is based on the Meuse River flood in December 1993, in the province of Limburg near the German and Belgian border. This was a river flood in an area that was protected by its natural elevation rather than dikes, which is uncommon in the Netherlands. Forecasting of water levels in the Meuse River was at that time relatively difficult, due to the limited ability to predict precipitation in the Meuse basin and the quick response of the river (Wind et al., 1999). Therefore the local authorities were caught by surprise, and there was insufficient time to warn the public (Wind et al., 1999).

The flood caused 254 million Guilder (price level 1993) in damages within the Netherlands, which is about 180 million Euros (price level 2016). The flood event inundated about 180 km^2 of land, which is 8% of the Province of Limburg. About 32% of the damage was to residential buildings and content, which are the damage categories considered in this study.

After the Meuse flood of 1993 the Dutch national government compensated the damage. The damage was collected by insurance experts working for a governmental organization called "Stichting Watersnood 1993". Shortly after the compensation the data was shared with WL Delft (now Deltares). The data was used in 1994 to create a flood damage model for the Meuse basin, in order to inform decision makers with respect to required flood protection works. WL Delft also applied several manual adjustments to the dataset (WL Delft, 1994) because the structural damage to rental houses was incomplete. What these adjustments were exactly is unknown. The only dataset now available is the dataset reported in the WL Delft study. More information about the background of these data and adjustments are reported in WL Delft (1994) and in Wagenaar et al. (2017).

The dataset contains 4398 complete damage records. Apart from the total flood damage, the dataset also contains the individual damages for structural building and content damage. Besides the damage variables the dataset also contains variables on water depth relative to floor level, household size, whether the house is detached and whether the house has a basement.

Wagenaar et al. (2017) added the following additional variables to this dataset, based on 2D flood simulations for the 1993 event: Water depth, flow velocity, flood duration and flood return period at the building location. Based on the Cadaster data the following variables were added: building age, building footprint area and building floor area. These variables were added based on the location of the building and there was significant uncertainty in this process because the building location was only available as a 6 digit postal code. More information about this is provided in Wagenaar et al. (2017). Table 4.1 shows the variables in the Dutch dataset.

4.2.1.3 Comparison and harmonization of variables

The compilations of the German and the Dutch datasets differed in respect to flood events and data acquisition methods. The German data was collected via surveys after six flood events, including the two most severe floods in Germany at least since 1950 (Schröter et al. 2015), while the Dutch data was collected by insurance experts during the compensation process of one flood event at the Meuse River. Thus, the German data covers a substantially larger variety of damage processes, including extreme situations. This is also reflected in the range and densities of the variable values, as shown for three examples in Figure 4.1. The variables in the German dataset show a considerably larger spread and often cover a larger range of values. For example, the German data contains a relatively high amount of cases with negative water depth (i.e. basement flooding only) but also with water depth above 200 cm (Figure 4.1). In contrast, most cases in the Dutch dataset have a water depth between zero and 200 cm. The German dataset also contains many more extreme values. Thus, also the mean relative building and content damages of the German dataset are higher in comparison with the Dutch dataset (Figure 4.1).

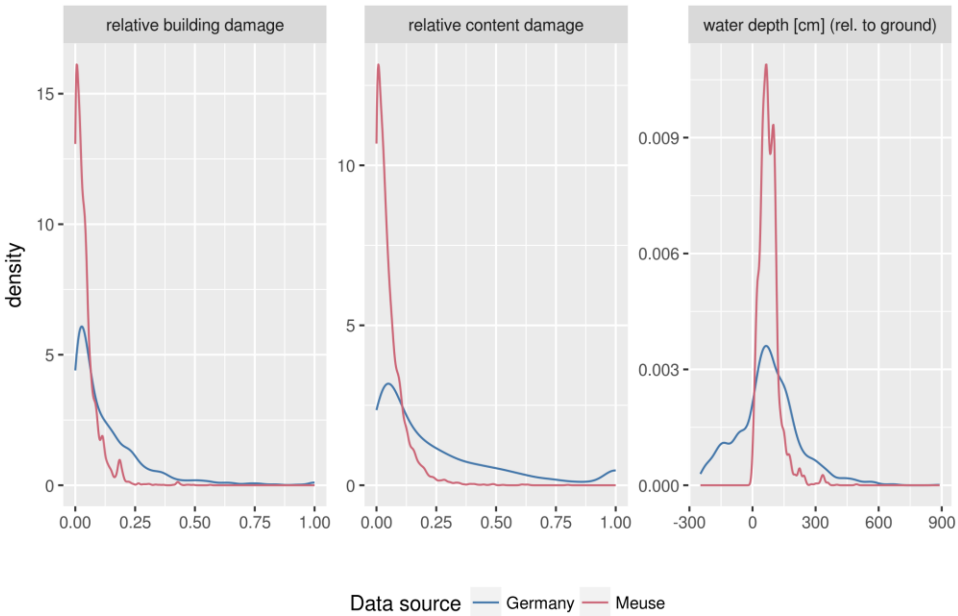


Figure 4.1: Comparison of the value ranges and densities of the relative building and content damage as well as water depths for the Dutch and German datasets.

Due to the different data acquisition methods, the definitions and units of the variables differ between the datasets. Thus, a harmonization of the variables was necessary to enable cross validation and updating procedures in which data of different datasets is mixed. Adjustments were made to the flow velocity and return period variables. The flow velocity in the German dataset is translated from its intensity class to a meter per second value to match the Dutch dataset. The return period definition of the Dutch dataset has been replaced with the German definition. In the German dataset the return period represents the return period of the flood that caused the damage, so it says something about the magnitude of the flood. In the Dutch dataset it is the flooding probability of any flood at that location. By switching to the German definition all records in the Dutch dataset get the same value, namely the return period of the 1993 event, which is 40 years. Table 4.1 shows an overview of the variables used in this study.

Table 4.1: Overview of the variables used in this study

<table>
<tr><th></th><th colspan="3">Variable</th></tr>
<tr><th>Abbreviation</th><th>Dutch dataset</th><th>German dataset</th><th>Unit</th></tr>
<tr><td>rsd</td><td colspan="2">Relative building damage[a,d]</td><td>Local currency</td></tr>
<tr><td>rcd</td><td colspan="2">Relative content damage[a,d]</td><td>Local currency</td></tr>
<tr><td>wdf</td><td colspan="2">Water depth relative to floor[a,d]</td><td>meter</td></tr>
<tr><td>bt</td><td colspan="2">Building type[a,d]</td><td></td></tr>
<tr><td>fa</td><td colspan="2">Footprint area of the building [c,d]</td><td>Square meter</td></tr>
<tr><td>wdt</td><td>Water depth relative to DEM[b]</td><td></td><td>cm</td></tr>
<tr><td>bs</td><td>Basement[a,d]</td><td></td><td>1=Yes, 0=No</td></tr>
<tr><td>hs</td><td>Household size[a,d]</td><td></td><td>number</td></tr>
<tr><td>fv</td><td>Flow velocity[b,d]</td><td></td><td>Estimated from score to m/s</td></tr>
<tr><td>ba</td><td>Building age[c,d]</td><td></td><td>Year</td></tr>
<tr><td>fal</td><td>Floor area for living [c,d]</td><td></td><td>Square meter</td></tr>
<tr><td>fd</td><td colspan="2">Flood duration[b,d]</td><td>Hour</td></tr>
<tr><td>rp</td><td colspan="2">Return period[b,d]</td><td>Year</td></tr>
<tr><td>fe</td><td></td><td>Flood experience[d]</td><td>Score</td></tr>
<tr><td>pre</td><td></td><td>Precautionary measures[d]</td><td>Score</td></tr>
</table>

[a] WL Delft, 1994 (Meuse)
[b] 2D flood simulation data using WAQUA (Meuse)
[c] Basisregistraties Adressen en Gebouwen (BAG), version 2011 (Kadaster website) (Meuse)
[d] Flood event surveys (Germany)

4.2.2 Damage models

4.2.2.1 Bayesian Networks based damage models

In Bayesian Networks, random variables and their conditional dependencies are represented in a directed acyclic graph (DAG) structure. Each variable can be either observed or represented as a prior probability distribution. Dependencies between variables are represented with edges representing joint probability distributions.

The edges in a Bayesian Network are directed which means there is a direction in which the influence of one variable flows to the other. From this network, the probability distributions of any variable can be predicted based on knowledge about the other variables.

The Bayesian Networks in this paper all use discretized variables. This means that each variable can take only a limited number of values. The advantage of this is that the joint probability distributions can be represented as tables (conditional probability tables) rather than as functions. The probability of a set of variable values can then be calculated with the following formula:

$$P(X_i, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{parents}(X_i))$$

Where X_i are the variables and $\mathrm{parents}(X_i)$ is the set of variables directed to X_i. The probability of a single variable value can be obtained by taking the sum of all the probabilities that contain the variable value of interest. $P(X_i|\mathrm{parents}(X_i))$ can be looked up from the conditional probability tables.

Conditional probability tables show for each combination of parent variable values the probability of each possible output value. The tables can be trained with data using maximum likelihood estimation. In this method the conditional probabilities are computed based on the number of observations that have the particular output value. The drawback of this method is that the number of possible combinations of parent variable values and output values can get very large. In such cases the number of observations is likely to become insufficient for the maximum likelihood estimation (Koller and Friedman, 2009). Discretization can therefore not become too fine and one variable shouldn't have too many edges pointing towards it.

German Bayesian Network based damage model (BN-FLEMOps)

The German Bayesian Network-based damage model is referred to as BN-FLEMOps, which stands for Bayesian Network Flood Damage Estimation MOdel for the private sector. The structure of BN-FLEMOps network is based on all complete records with respect to the variables water depth, relative damage, return period, flood duration, building area, building type, precautionary measures, and flood experience from the German database described in section 2.1.1. A set of three different algorithms implemented in the R-package "bnlearn" (Scutari, 2010) were used in a bootstrap approach to learn the network structure shown in Figure 2. With a random sample of 950 observations (out of 1456) the algorithms Fast-IAMB (Fast Incremental Association (Yaramakala and Margaritis, 2005), Inter-IAMB (Interleaved Incremental Association (Tsamardinos et al., 2003) and a hill-climbing approach using the Bayesian Dirichlet Equivalent (BDE) (Heckerman et al., 1995) were initialized 500 times. The result set of 1500 network structures in total provided the basis to define the connections between the variables. All arcs, and their associated directions, that occurred in at least 80 percent of the cases within the result set have been used to support the development of the expert network BN-FLEMOps. The bootstrap approach that uses only the subset of 950 observations was used to avoid over fitting and to ensure a robust network structure.The variables were discretized on an equal frequency basis with water depth and relative damage in 10 classes, return period and duration in 5 classes and building area in 3 classes. The other variables are discrete by definition.

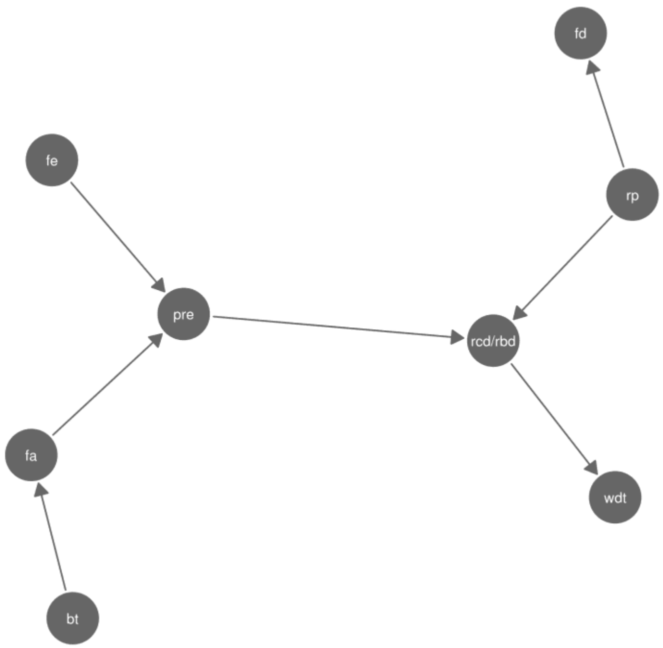


Figure 4.2: Structure of the German Bayesian Network based Flood Damage Estimation MOdel for the private sector BN-FLEMOps

Dutch Bayesian Network (Meuse-BN)

In the Dutch Bayesian Network, Meuse-BN, both the content and building damage are included in the same network, and the network is used to determine both values simultaneously (conditional on the other variables in the network). The Meuse-BN model works with the libpgm Python library (Cabot, 2012), which is based on the methodology described in Koller and Friedman (2009). The discretization is based on the principle of having an equal number of observations per bin, because this is an efficient way of using the limited number of observations. The number of bins per variable value is determined by testing several configurations and then picking the one with the smallest mean absolute error (MAE). The result is 6 bins for the output variables, 4 bins for the water depth variables, 3 bins for the living area and other hydraulic variables and 2 bins for the rest of the variables. More details about the Meuse-BN model can be found in Wagenaar et al. (2017).

The network structure of the Dutch Bayesian Network was learned on the basis of all available 4398 records from the Dutch dataset and the predefined discretization. A hill climbing approach together with the BDE (Heckerman et al., 1995) was used to find the structure with the highest BDE score. The approach uses a sequence of arc additions, arc removals and arc inversions to maximize the BDE score, initially starting with no arc between the variables at all. This approach was repeated 1000 times to avoid local minima in the search space. The structure of the resulting Bayesian network is shown in Figure 4.3.

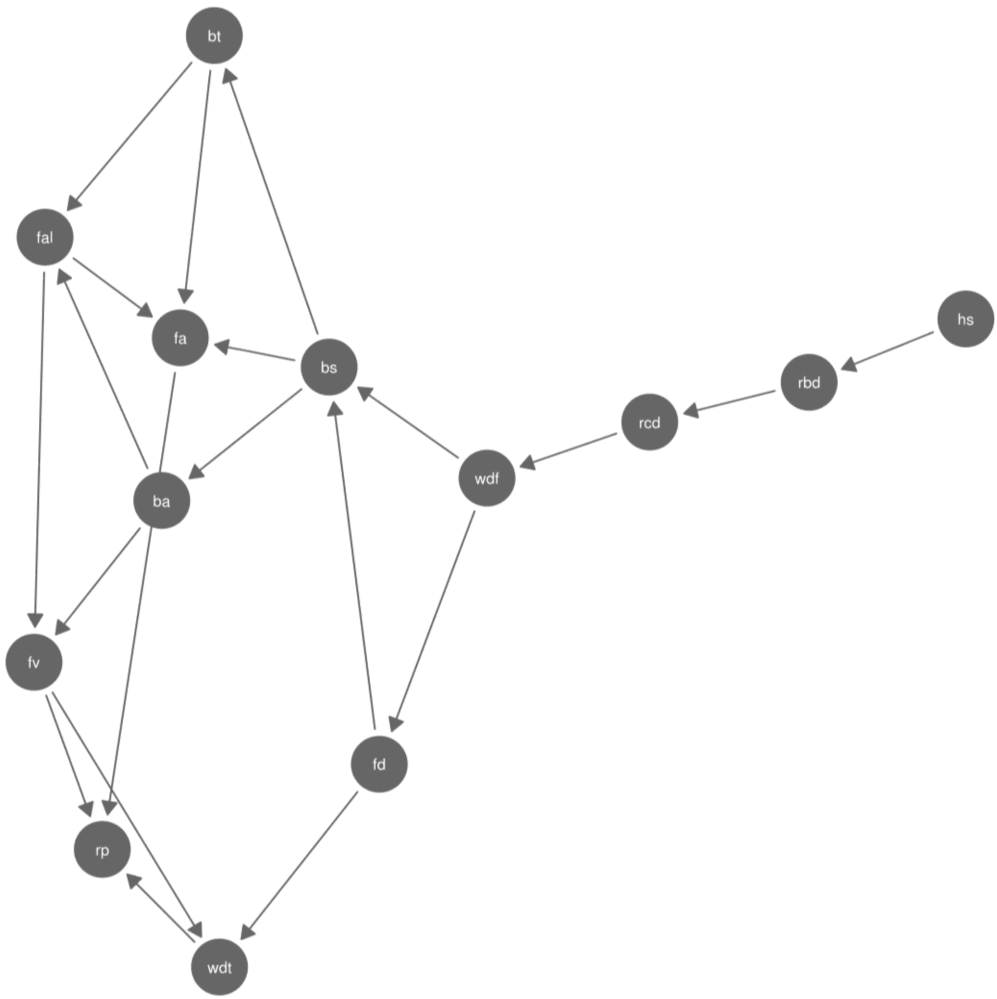


Figure 4.3: Structure of the Dutch Bayesian Network based model Meuse-BN.

4.2.2.2 Random Forest-based damage models

Random Forests are ensembles of regression trees. A regression tree is a series of binary question nodes (e.g. water depth>0.5m) about the input variables leading to an output value of the target variable (i.e. relative building damage or relative content damage). These trees are typically created with regression tree learning based on a dataset with observations.

A regression tree is a serie of questions about the data (e.g. water depth>0.5m?) on which the data is split into different branches forming a tree of options. This tree can be seen as a model to make new predictions. During regression tree learning, an algorithm is searching for variables and values within the data set on which to split the data in order to maximize improvement in homogeneity of the output variable value observations. After each splitting question the tree is further build up and the resulting branches with observations get more homogenous. A central question of regression tree learning is when to stop splitting up the observations. If the tree gets too complex, it will perform poorly on predicting new observations. Several methods are available to stop the tree from growing too complex.

One of the advantages of a random forest is that it helps to prevent overfitting and hence improve the predictive capacity of the model. In a random forest a large number of regression trees are built based on resampling the observations. Furthermore, at each split of each tree in a random forest a number of variables are left out as potential splitting candidates. The average of the predictions of all trees is the prediction of the entire random forest. The variation among the trees represents the probability distribution of the prediction.

German Random Forest based damage model (RF-FLEMOps)

The models for the German Random Forest based Flood Damage Estimation MOdel for the private sector (RF-FLEMOps) were learned using the R-package "randomForest" (Liaw and Wiener, 2002). A sensitivity analysis was conducted to find the most robust configuration for the number of trees and the number of variables that were considered at each node. The sensitivity analysis was done visually,to evaluate the convergence of the "out of bag error.". The "out of bag" error is the prediction error of an independent part of the training data that is not used to build an individual tree but instead used to validate the performance. Convergence was reached with 5000 trees and the default value using 1/3 of the variables at each node for splitting.

Dutch Random Forest (Meuse-RF)

The Random Forest trained on the Meuse data (Meuse-RF) works with the Python library "Sci-Kit learn". Overfitting is prevented by limiting the number of splits in each tree to 25, the value which resulted in the smallest MAE on validation data. During the random forest learning 2/3 of the variables are randomly left out as potential splitting candidates at each split. This is a standard Sci-Kit learn configuration for Random Forests. One hundred different regression trees are built in the random Forest.

4.2.3 Cross-country model transfer setup

4.2.3.1 Sampling of training and validation datasets

The German models BN-FLEMOps (for building damage) and RF-FLEMOps (for content damage) were initially trained on 350 randomly selected German data points. Similarly, the Dutch models Meuse-BN and Meuse-RF were initially trained on 350 Dutch data points. The German and Dutch models were then validated twice using 200 data points sampled from both the German and Dutch datasets. This set-up is shown in figure 4.4. The training and validation process was repeated 1000 times, resulting in a distribution of model performance results.

Some input variables of the damage models BN-FLEMOps and RF-FLEMOps have no (near) equivalent in the Dutch dataset, such as the precautionary measures and flood experience indicators. Missing variables in the validation dataset increase the uncertainty and decrease the probability that a damage observation is similar to the modelled damage value. To address the problem of missing values in Bayesian Networks, which can calculate with any number of unknown variables, the missing data is treated as an extra unknown variable (together with the relative damage). The BN calculates a posterior distribution for the missing variable (conditional on the remaining variables), from which a value is randomly sampled. For the Random Forests, no assumptions about missing data are made and the missing values are filled by drawing from a uniform distribution that is defined by the minimum and maximum of the original variables with no missing values.

4.2.3.2 Updating models with different datasets

In another analysis all the Dutch (Meuse-BN, Meuse-RF) and the German (BN-FLEMOps, RF-FLEMOps) models are in a first step updated with 350 randomly selected extra data points from the dataset of the other country and in another second step again updated with additionally 350 data points of the dataset of the other country. The models are then again validated after the first and the second updating step based on 200 independently randomly selected validation data points from the dataset of the other country. This process is also repeated 1000 times to get rid of any randomness in the sampling. The hypothesis is that the models will perform better in this cross-country transferability test when they have been updated with (independent) data from the country for which they are validated on.

Since some input variables of the damage models BN-FLEMOps and RF-FLEMOps have no equivalent in the Dutch dataset, dealing with missing values during the updating procedure is necessary. In case of the model BN-FLEMOps, missing data was filled following the joint probability distribution of the model without the new data. This approach ensures that existing dependencies between variables defined by the Bayesian network structure are considered and applied. Thus, existing knowledge about damage processes is transferred during the updating process. In case of the model RF-FLEMOps, a linkage between variables does not exist in such an explicit structure, but is rather hidden under the total number of trees in the Random Forest. Because this structure is hard to address, no assumptions about missing data are taken and the missing values are filled by drawing from a uniform distribution that is defined by the minimum and maximum of the original variables with no missing values.

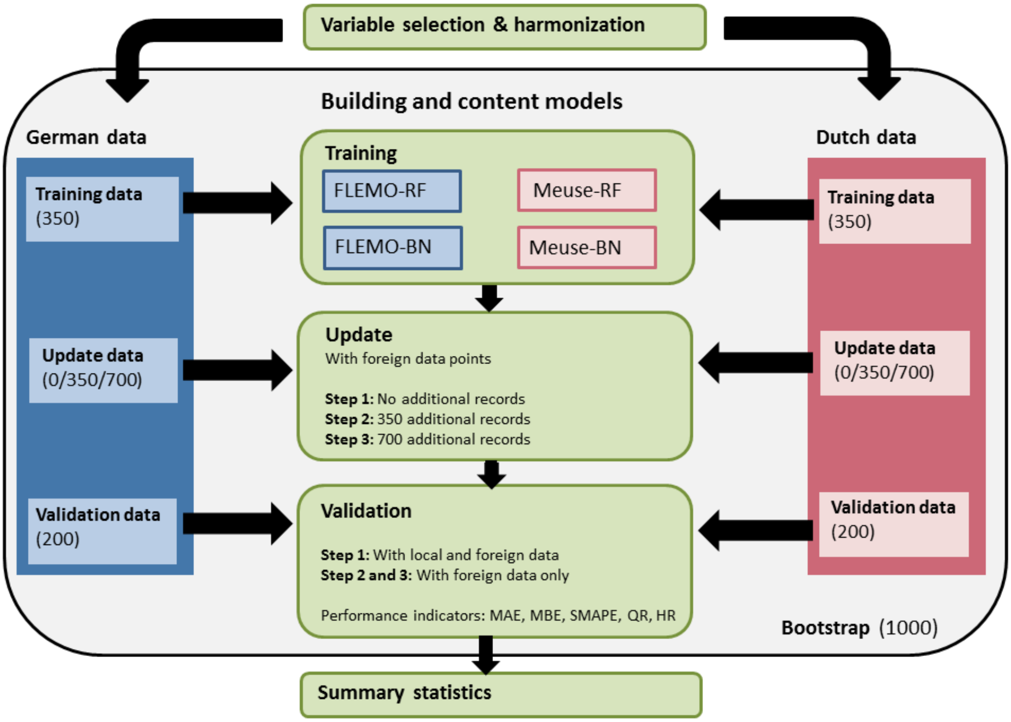


Figure 4.4: Overview of the cross-country model transfer set-up

4.2.3.3 Evaluation criteria

To evaluate the performance of the models 5 different evaluation criteria are applied (see table 2). These are: Mean Absolute Error (MAE), Mean Bias Error (MBE), Symmetric Mean Absolute Percentage Error (SMAPE), Hit Rate (HR) and Quantile Range (QR). Both, Bayesian Networks and Random Forests provide multiple estimations for every individual validation point or prediction. For Bayesian

Networks, this is the result of conditional probabilities, for the Random Forests, it is given by the single prediction of each tree. MAE, MBE and SMAPE evaluate the precision of the models using the median of the simulations for every individual validation point. The MAE shows the mean error among all observations used for the validation. The errors are made absolute before taking the mean and therefore an overestimation cannot be used to balance out an underestimation. This error metric is especially useful when one is interested in the model performance on individual objects, which is of interest for insurance purposes, for instance. This is also an absolute error metric in the sense that the error is expressed in relative damage. Therefore, a large error on a large damage is equal to the same large error on a small damage. The SMAPE error metric corrects for this by calculating a relative error. The sum of the absolute errors is divided by the sum of all observed and simulated damages. This means that a large error on a small damage is more important than the same large error on a large damage. This is relevant here because the average Dutch damages are much lower than the German damages (see 2.1.4).

The MBE is an error metric in which overestimates are allowed to be compensated by underestimates. This is relevant when the purpose of the model is to estimate total damage such as often is the case in studies related to public policymaking on flood risk, and for societal benefit-cost analysis.

The quality of the simulations is further evaluated by the QR (Quantile Range) and the HR (Hit Rate). The QR evaluates the variation within the model predictions and the HR the reliability of the models. The QR is the distance between the 5% and 95% quantiles of the estimated flood damage. A large QR means the uncertainty of the prediction is high according to the model. The HR is the fraction of the observations that lie within this 90% QR. A HR of 0.9 means a perfect reliability (Thordarson et al., 2010). A HR larger than 0.9 indicates that the QR is probably too large (10% of the observations should be outside the 90% QR). A smaller HR shows that the QR is too small. All models are trained and validated with 1000 different sample combinations, which allows the construction of a distribution of the observed quantile ranges and hit rates.

Tabel 2. Formulas of the error metrics. $RL_{sim,n}$ is the nth simulated relative damage, $RL_{obs,n}$ is the nth observed simulated damage. N is the total number of observations.

Error metric	*Formula*
Mean Absolute Error (MAE)	$MAE = \frac{1}{N}\sum\left\|RL_{sim,n} - RL_{obs,n}\right\|$
Mean Bias Error (MBE)	$MBE = \frac{1}{N}\sum RL_{sim,n} - RL_{obs,n}$
Symmetric Mean Absolute Percentage Error (SMAPE)	$SMAPE = \frac{\sum\left\|RL_{sim,n} - RL_{obs,n}\right\|}{\sum\left\|RL_{sim,n}\right\| + \left\|RL_{obs,n}\right\|}$
Quantile Range (QR)	$QR = \frac{1}{N}\sum RL_{sim,n,q95} - RL_{sim,n,q5}$
Hit Rate (HR)	$HR = \frac{Observations\ within\ QR}{N}$

4.3 Results and discussion

In this section first the results are shown from the cross-country model transfers. These results are compared with the model performance on local data and implications of these results are discussed. Then the results are shown from the models updated with data from the country it is applied in and the improvement in results is shown and are again discussed.

4.3.1 Initial model performances

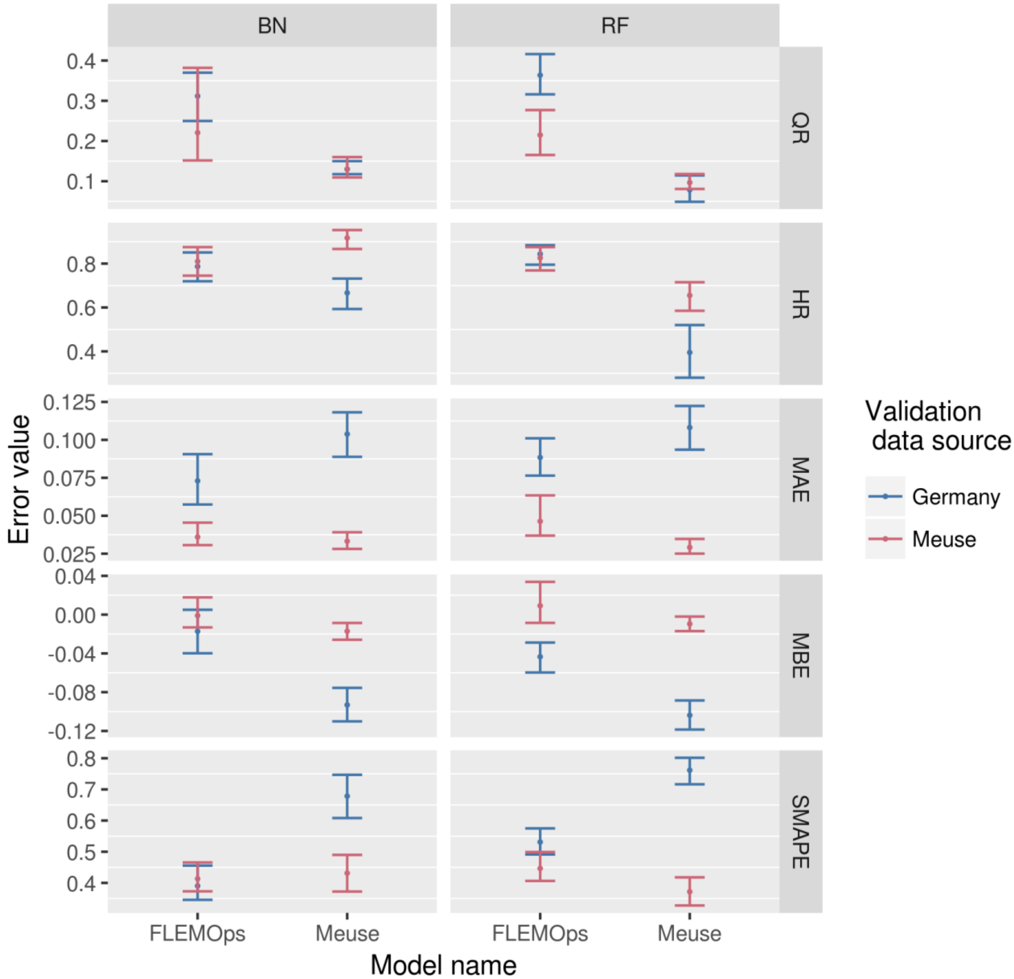


Figure 4.5 Validation results for building damage. Left the results for the Bayesian Networks and right the results for the Random Forests.

The models BN-FLEMOps and RF-FLEMOps, which were trained on the German data predict the Dutch damages much better than the other way around (i.e. the Meuse models predicting the German damage). This is consistent over all precision error metrics (MAE, MBE and SMAPE). In a few instances the BN-FLEMOps even outperforms (by a small but significant margin) the Meuse-BN in estimating the Dutch damage (SMAPE and MBE error metrics in Figure 4.5 and 4.6). The probable reason for this is that the German dataset is much more heterogeneous than the

Dutch dataset. The cause of the larger heterogeneity is the mix of various flood events and large, diverse regions from which data is included. The Elbe catchment (former East Germany) and Danube catchment (Bavaria) have very different building stock and socio-economic characteristics. Some of the observations are therefore possibly representative of the Dutch situation. This allows the FLEMOps models to do well in the Netherlands because the spectrum of flood damage data that it was trained on is much larger. The Dutch data comes from only one flood event in a single relatively small region, with relatively homogenous building stock. Thus, the damages and other variables are more homogenous and the size of the spectrum of its observations are smaller. Therefore, there will be many observations in the German data outside the small spectrum of the Dutch data and hence the Meuse models need to extrapolate when trying to predict these observations. Most Dutch damage observations will be within the large spectrum of German observations and hence, the FLEMOps models mainly only need to interpolate when estimating Dutch damages.

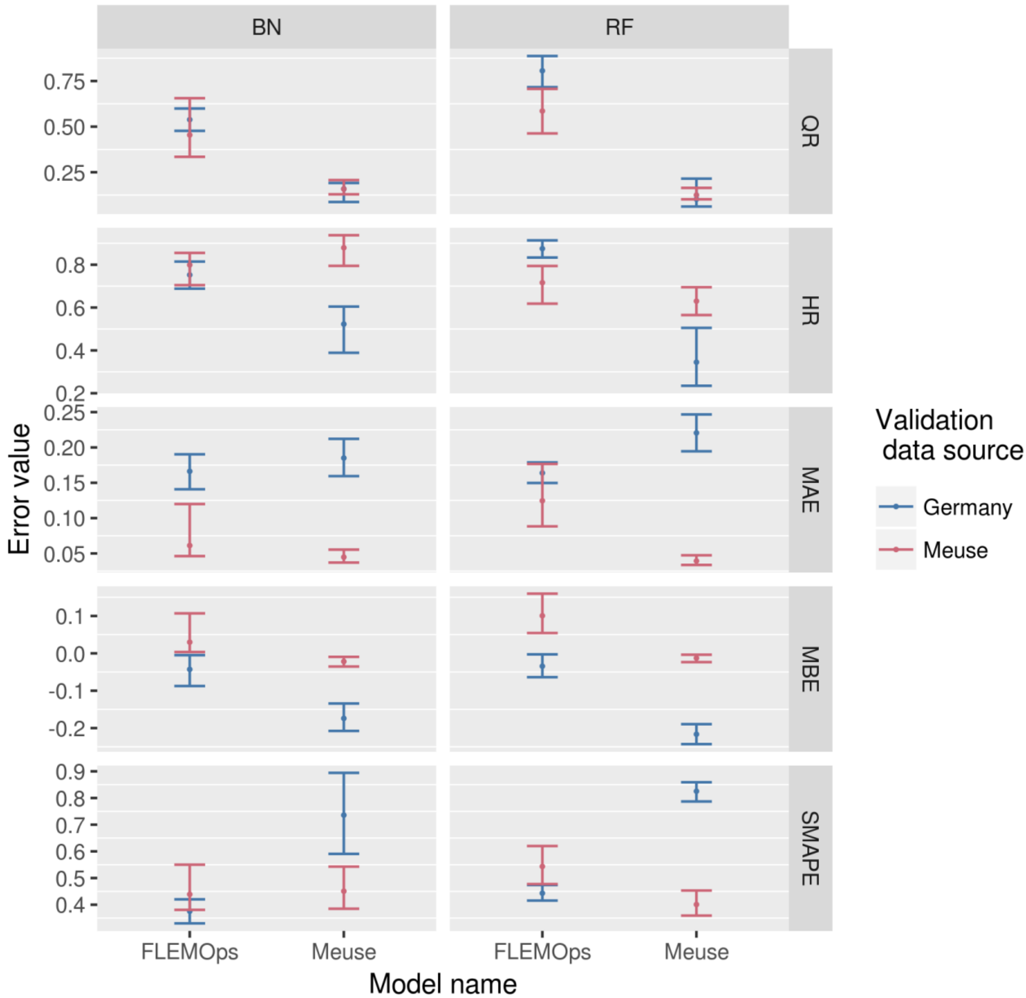


Figure 4.6 Validation results for content damage. Left the results for the Bayesian Networks and right the results for the Random Forests.

The Dutch dataset has a larger number of data points than the German dataset, 4398 (building and content) in comparison with 1456 (building) and 1324 (content) , respectively. Because of the sampling setup in this study this potential advantage is not used in the current study. However, several tests with taking more data points for training have shown no significant improvements when sampling a larger number of data points for training. It is therefore expected that extra data points are only valuable when they add to the spectrum of the model. This is more evidence, that when aiming on development of an empirical flood damage model, for flood damage data collection the heterogeneity of the collected data, and thus the reflection of different damage processes, is more important than the size of the dataset.

The Quantile Range results show a similar story. The Meuse models are trained on relatively homogenous data and are therefore much more confident about their predictions, which is reflected in a small QR. The consequence however is a poor HR of the Meuse models validated on the German data. While Meuse-BN almost scores a perfect HR of 0.9 when validated on the Dutch data, reflecting a high reliability, the FLEMOps models have good HRs around 0.8, on both the Dutch and German data.

A notable result is that FLEMOps performs for the MAE significantly better in predicting the Dutch damage than in predicting the German damage. This is unexpected because a model is believed to perform better in the area its training data originates from. This is probably because the damages in the Netherlands are on average much lower than in Germany (see 2.1.4) and therefore the absolute errors are also smaller. The relative error metric corrects for this and hence on the SMAPE error metric the FLEMOps models work better when predicting the German data 3 out of 4 times. Similar observations are given for the comparison between the estimates for the building and contents damage. MAEs are significantly larger for estimating relative contents damage in comparison with relative building damage for all models. Such a significant difference cannot be observed for the SMAPE. This shows that when comparing performances from different validation sets a relative error metric adds additional information and thus should complement the absolute error metric.

In the cross country validation test, i.e. FLEMOps validated with Dutch data and Meuse model validated with German data, the Bayesian Network based models perform better in comparison with the random forest based models.

4.3.2 Model performance after update

The updating (see figure 4.7 and 4.8) test confirms our hypothesis: All models improve their performance when validated on the dataset of the other country after they are updated with independent data from that country. This improvement is visible on almost all error metrics. The MBE for building damage is an exception here, the performance of the FLEMOps declines slightly (but significant) after an update with Dutch data (validated on Dutch data).

The improvement of model performance after updating is high for the Meuse models when updated and validated with German data. The performance improvement of the FLEMOps models is smaller in comparison with the Meuse models.

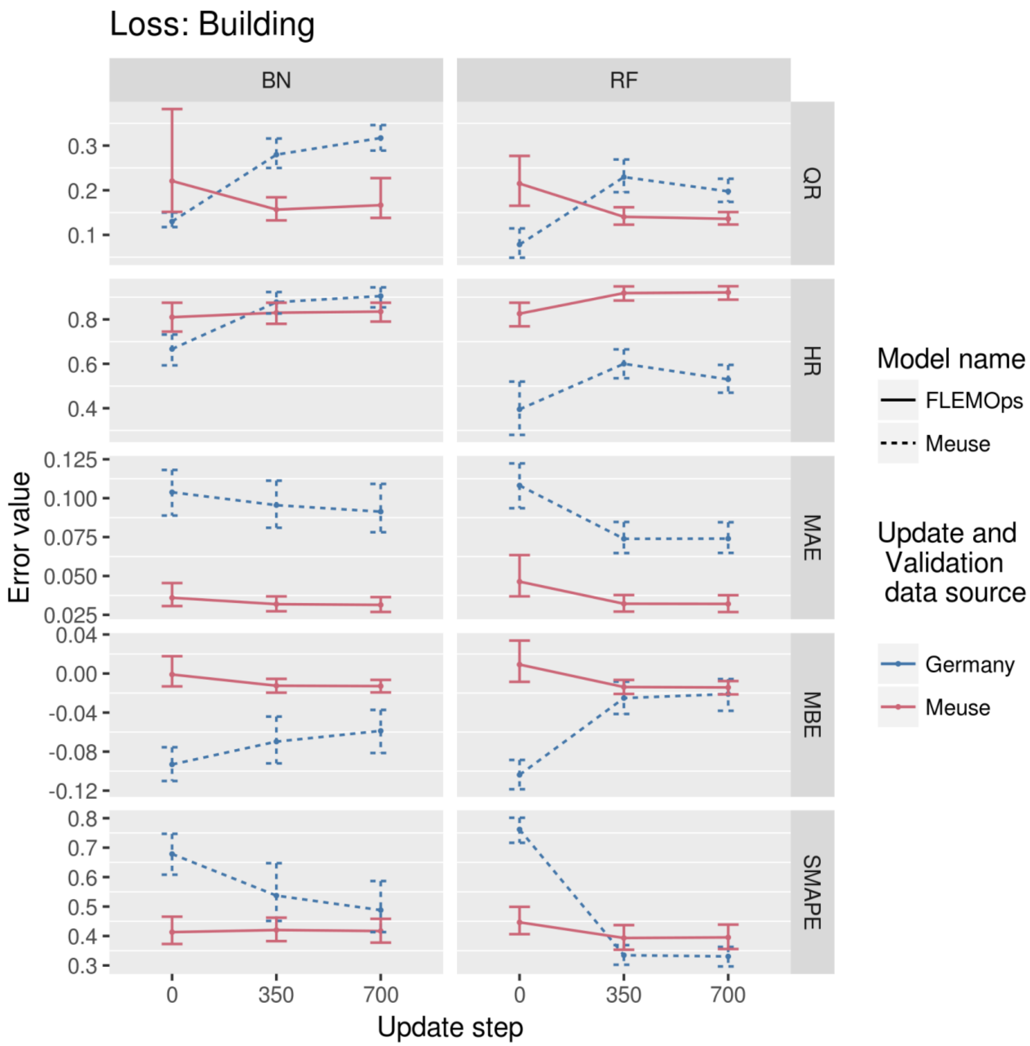


Figure 4.7 Validation results for building damage after update with foreign data. Left the results for the Bayesian Networks and right the results for the Random Forests.

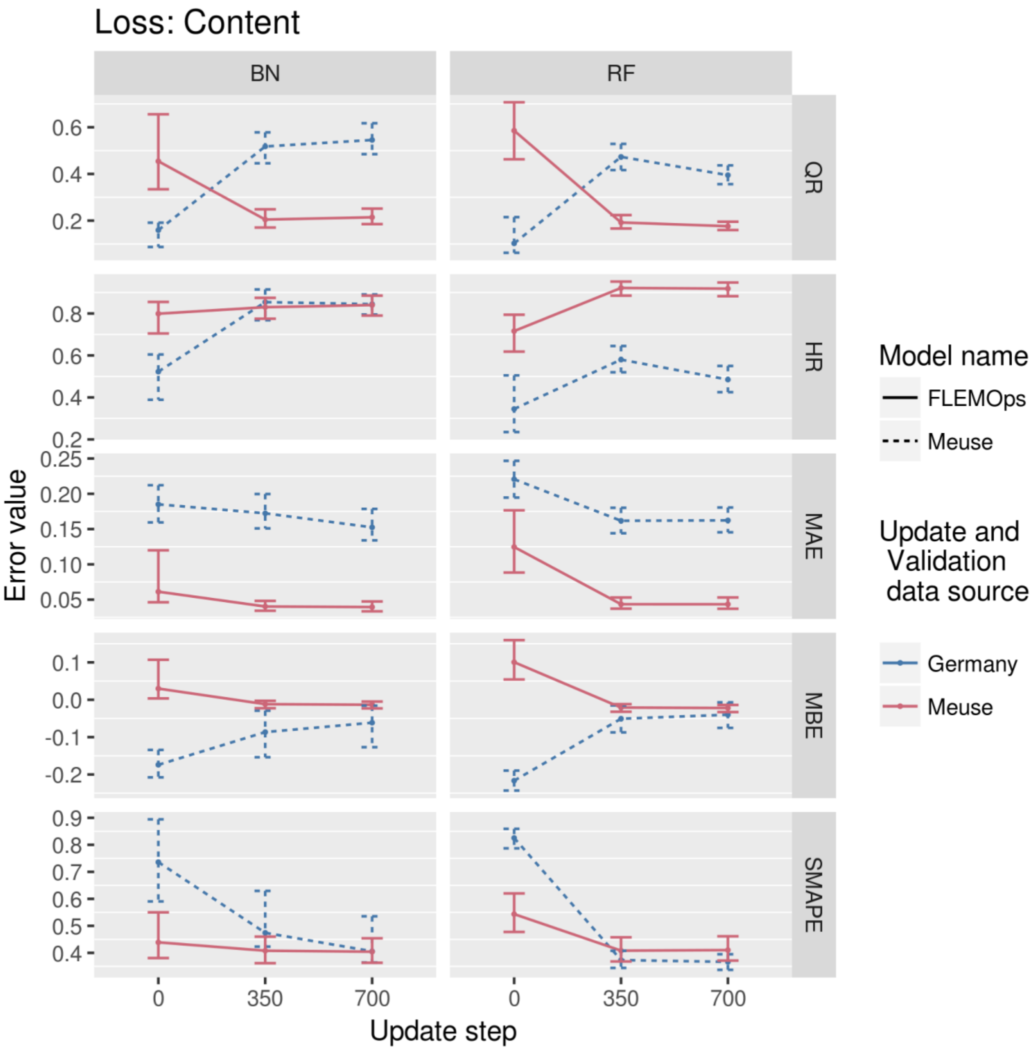


Figure 4.8. Validation results for content damage after update with foreign data. Left the results for the Bayesian Networks and right the results for the Random Forests.

For instance, updating BN-FLEMOps hardly leads to improvements for estimating building damage, however, it does for estimating content damage. This can also be explained based on the homogeneity of the Dutch data. Because the Dutch data is homogenous it profits a lot by extra data from other events, while the German data already has data from many events and therefore profits less by the data of another event, even if this data comes from the same source as the validation dataset.

Generally, the performances of the Bayesian Network models does not differ much from the ones of the Random forest models. However, during the updating procedure, the Random forest models seem to profit more from additional data than the Bayesian Network models. This is as expected because updating the Bayesian Network just updates the conditional probability tables, not the relation between the variables (i.e. the Bayesian Network structure). Thus newly introduced damage processes are not fully reflected by the updated models. For Random Forests on the other hand, the update learns an entire set of new trees and thus damaging processes that are given by the new data at hand can be described by the updated models.

The bootstrap approach assesses the sensitivity of the data selection for model training and validation. As the German data has a higher variability, the FLEMOps models trained on that data are sensitive to the selection because data might be chosen that is similar to the Dutch validation data and thus show low errors, or data might be chosen that is very different to the Dutch validation data and thus show higher errors. This would have been a smaller issue if more than 350 data points would be used for the zero step. As a consequence, the spread for the error measures at step 0 is bigger than at the following steps since the updated data used for the next steps is from the same population as the validation data, and as we add data from the same source as the validation set, we are more likely drawing data that is similar to the validation data. Thus the spread of the errors of the FLEMOps models is generally decreasing during the updating procedure. The Meuse models don’t have this issue because the data is homogenous and therefore the performance depends less on what data is sampled (it performs consistently bad on the German validation data).

4.4 Conclusions

The multi-variable flood damage models BN- and RF-FLEMOps perform better when validated on data from The Netherlands than the Meuse models when validated on German data. This is probably because the German training data is more heterogeneous, since they are based on a large number of events from different regions. The FLEMOps models, therefore, seem to be better able to capture damage processes from other events and in other locations. The Dutch data is only based on a single flood event in a relatively small area. This also explains why the Meuse models benefit more from updating with data from Germany, than the other way around.

This implies that the heterogeneity of collected flood damage data for modelling purposes can be more important than the quantity of data points. Future flood damage collection efforts should therefore focus on acquiring heterogeneous data, for instance by targeting different locations in the same event, or several events for the same location, instead of collecting a large quantity of data only for a single event in one location. Such data collection efforts over several years and events may provide much more relevant information for improving risk analysis than an in-depth study for a single event.

The good performances of the FLEMOps models in the Netherlands show that the transfer of multi-variable flood damage models in space and time can at least in some cases be successful. The important condition for this seems to be that the training data contains observations collected under similar conditions as present at the location to which the model is transferred. Performance of models in transfer settings is additionally improved via updating the models using data from the location in which they are applied.

Acknowledgments, Samples, and Data

This research was supported by the European Union’s Horizon 2020 research and innovation programme, through the IMPREX project (Grant Agreement number 641811). Parts of this work have been supported by funding for the OASIS+ Future Danube: Multi Hazard and Risk Model demonstrator project through the Climate Knowledge and Innovation Community (Climate-KIC) by the European Institute of Innovation and Technology (EIT) and from the European Union's Horizon 2020 research and innovation program (Grant Agreement No 730381). The surveys to collect the empirical damage data in Germany were supported by the German Research Network Natural Disasters (German Ministry of Education and Research (BMBF), no. 01SFR9969/5), the MEDIS project (BMBF; no. 0330688) the project “Hochwasser 2013” (BMBF; 13N13017) and by a joint venture between the German Research Centre for Geosciences GFZ, the University of Potsdam and the Deutsche Rückversicherung AG, Düsseldorf.

We would like to thank the editor Jim Hall and two anonymous reviewers for their constructive and useful suggestions. We would also like to thank Kathryn Roscoe for proof-reading the manuscript.

5. SAMPLE SELECTION BIAS CORRECTION

This chapter is based on the paper "Improved transferability of data-driven damage models through sample selection bias correction" written together with Tiaravanni Hermawan (Deltares), Marc van de Homberg (510 Red Cross), Jeroen C.J.H. Aerts (VU Amsterdam), Heidi Kreibich (GFZ), Hans de Moel (VU Amsterdam) and Laurens Bouwer (Climate Service Center Germany). It is currently under review for publication.

Abstract

Damage models for natural hazards are used for decision-making on reducing and transferring risk. The damage estimates from these models depend on many hazard, exposure and vulnerability variables and their complex non-linear relationships with the damage. In recent years, data-driven modeling techniques have been used to capture those relationships. The available data to build such models is often limited. Therefore, in practice it is usually necessary to transfer models to a different situation. In this paper we show that this implies that the samples used to build the model are often not fully representative for the situation where they need to be applied on, which leads to a 'sample selection bias'. In this paper, we enhance data-driven damage models by applying methods, not previously applied to damage modelling, to correct for this bias before the Machine Learning models are trained. We demonstrate this with case studies on flooding in Europe, and typhoon wind damage in the Philippines. Two sample selection bias correction methods from the machine learning literature are applied and one of these methods is also adjusted to our problem. These three methods are combined with stochastic generation of synthetic damage data. We demonstrate that for both case studies, the sample selection bias correction techniques reduce model errors, especially for the mean bias error this reduction can be larger than 30%. The novel combination with stochastic data generation seems to enhance these techniques. This shows that sample selection bias correction methods are beneficial for damage model transfer.

5.1 Introduction

Over the last decades, both the developed and the developing world have seen an increase in the frequency and severity of hydro-meteorological disasters and their impacts (Munich Re NatCatService, 2017). Many sectors are affected and can benefit from improved models to predict these impacts, so that better decisions can be taken to reduce, retain, transfer or absorb the risk (Van den Homberg & McQuistan, 2019). Natural hazard damage models predict the damages of a disaster given hazard characteristics such as the water depth of a flood (e.g. Merz et al., 2010) or the wind speed of a cyclone (Pielke, 2007). They are used to estimate risk from natural hazards in order to support decisions about investments in risk reduction measures. An example is their crucial role for determining the required protection levels of the dikes in the Netherlands (e.g. Kind, 2013; Van der Most, 2014). Damage models are also increasingly used for providing impact information in early warning systems (e.g. Bachmann et al., 2016), and many national meteorological and hydrological organizations are attempting to move from hazard forecasts to impact-based forecasts (WMO, 2015) whereby damage models are essential. Several actors,such as humanitarian organizations, can use these impact-based forecasts to initiate early actions that reduce risks just before a hazardous event (Coughlan de Perez et al., 2015). Once the disaster has hit, similar models can be used to prioritize humanitarian aid (risk absorption) (van Lint, 2015; van den Homberg et al., 2017; van der Veen, 2016). Damage models or so-called catastrophe models are also applied in the insurance sector to determine premiums (Pielke et al., 1999; Grossi and Kunreuther, 2005; Merz et al., 2010).

Traditionally, damage models often follow relatively simple approaches to estimate damages. For example, flood damage models typically relate the single parameter ‘water depth’ to resulting damage using ‘depth-damage curves’ (Merz et al., 2010). Whereas typhoon damage models similarly relate maximum wind speed to storm damage (van Lint, 2015; van den Homberg et al., 2017; Pielke, 2007). However, these simple models show considerable uncertainty in their damage estimates (de Moel et al., 2014; Wagenaar et al., 2016; Gerl et al., 2016), and do not always perform well when they are transferred (e.g. Jongman et al., 2012),. The main reason for the uncertainty is that the damage functions contain implicit assumptions about variables and processes that are not included in the model (Wagenaar et al., 2016). Examples of such variables are: Flood duration, flow velocity, building construction and materials, precautionary measures, contamination of the flood water and household size.

Nateghi et al., (2011) introduced machine learning (ML) methods to predict impacts of natural hazards (electricity outages from storms). Merz et al (2013) used a similar approach to predict flood damages at individual building level. Since then such techniques have been applied by many authors to predict all sorts of impacts from natural hazards: Nateghi et al., 2014, Schröter et al., 2014, Schröter et al., 2018, Sieg et al., 2017, Wagenaar et al. 2017, Mayfield et al, 2018, Carvajal et al., 2018, Wagenaar et al., 2018, Amadio et al., 2019 and Ganguly et al., 2019. These data-driven damage models often use more than one variable to predict the damage (multi-variable models). Therefore, they often perform better than traditional flood damage models (Wagenaar et al. 2017; Kreibich et al. 2017), particularly when models are transferred (Schröter et al., 2014; Wagenaar et al., 2018). In practice damage models are always applied in a transfer setting (Cammerer et al. 2013). This is for example a model built on data or knowledge from one location applied in another location (spatial transfer), or data collected at one moment in time being applied at a different time (temporal transfer). Detailed data on flood damages are rarely recorded in a structured and consistent way and are often outdated. Some recent examples where empirical damage data was collected are described by Kienzler et al. (2015), Poussin et al. (2014), and Molinari et al. (2014) for cases in Germany, France and Italy, respectively.

Machine Learning (ML) methods assume that the training data to build the model consist of randomly drawn samples from the same distribution as the test samples for which the learned model needs to make predictions (Zadrozny, 2004; Pan & Yang, 2010). In a spatial and temporal transfer setting this is often not the case. For example, damages from moderate typhoons may be used to predict the damage of an extreme typhoon. In such cases the ML algorithms need to rely on outlier observations in the data to build the most crucial part of the model. This problem is called the 'sample selection bias'. This received considerable attention in econometrics for the application to linear regression (Zadrozny, 2004). In the year 2000, Heckman (1979) received the Nobel prize in economics for developing a correction method. This "Heckman" correction however only applies to linear regression models. Cortes et al. (2008) provided two techniques to correct for this problem in case other ML methods are applied: these techniques are Cluster-Based Estimation (CBE) and Kernel Mean Matching (KMM). In this paper we apply, to our knowledge for the first time, sample selection bias correction techniques to damage models for natural hazards, and show their potential benefits. We also introduce a variation of the Cluster Based Estimation method that we call Single Variable Distribution Matching (SVDM), which only uses the most relevant variable.

Sample selection bias correction techniques give weights to the training data to make the most relevant samples more important during the training of the ML models. However, such techniques can result in very high weights given to single observations. In our analyses, we therefore explore a new combination of techniques where very high weights are smoothed out before they are included in the ML model. This is done by resampling the data after the sample selection bias correction with a statistical model. The resulting synthetic data is used to then train the machine learning models. This synthetic data generation in combination with sample selection bias correction methods is a new approach.

Sample selection bias correction techniques have never been applied to correct multi-variable data-driven models to predict the impacts of natural hazards. The objective of this research is therefore to evaluate how three sample selection correction techniques (CBE, KMM, SVDM) reduce the sample selection bias for multi-variable data-driven damage models and improve their performance when they are transferred between different events and between different geographic locations. These methods are evaluated with and without resampling of synthetic data and with two different machine learning methods: Artificial Neural Networks (Breiman, 2001) and Random Forests (Rumelhart et al., 1986). In total 12 different model setups are compared. These methods are applied to two different case studies where data-driven multi-variable damage models are transferred in time and space. The first case study is based on a dataset with typhoon damages in the Philippines on macro level (municipalities). The second case study is an extension of the paper of Wagenaar et al. (2018), where multi-variable micro-scale (buildings) flood damage models are transferred between the Netherlands and Germany. This paper starts with an explanation of the methods, including an introduction to the case studies, the data and the evaluation metrics used to assess the model performance. Next, the results are presented and discussed, and finally the conclusions are presented.

5.2 Methods and data

Figure 5.1 presents our method with six steps to improve damage estimation in transfer settings with data-driven multi-variable models based on Machine Learning techniques. The first step consists of selecting and developing training data for the damage models. This data comes from different events than the application (test) data for which the model needs to predict the damages. The second step is to apply three different sample selection bias correction techniques on a training dataset. The corrected training data is subsequently either directly used in two machine learning techniques (random forests and artificial neural networks) to estimate damages (step 4 and 5), or is first resampled using a statistical model (step 3). Step 3 is only applied to test the influence of generating synthetic data. The resulting damage estimates are evaluated with various error metrics (Mean Absolute Error, Mean Bias Error, Symmetric Mean Absolute Percentage Error)(step 6). This approach is applied to both case studies (flood damage and typhoon wind damage). Below, the data-driven approaches are further described (section 5.2.1), the case studies are introduced (5.2.2), the specific model setup to apply the data driven approaches to the case studies is shown (5.2.3), and finally the evaluation metrics are specified (5.2.4).

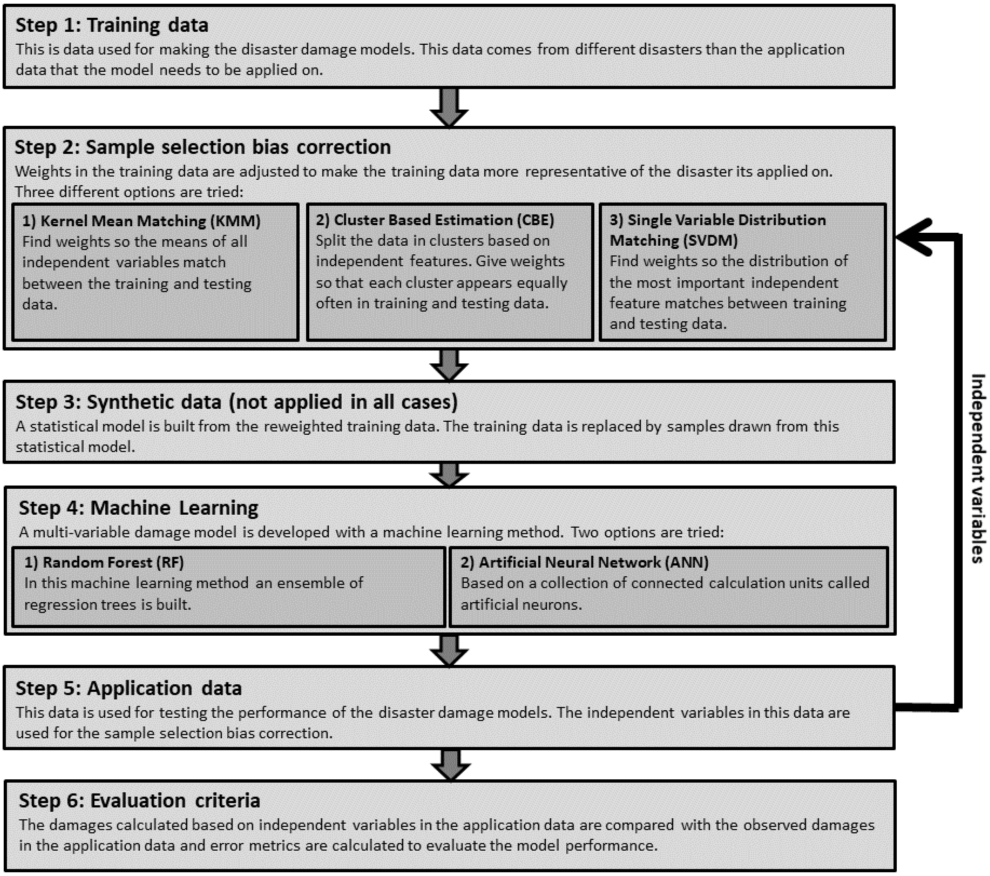


Figure 5.1: Overview of the approach for developing multi-variable damage models from observational data, including the testing procedure.

5.2.1 Data-driven methods

5.2.1.1 Sample selection bias correction

Kernel Mean Matching (KMM)

Kernel Mean Matching (KMM) (Cortes et al., 2008) assigns a set of weights to the training data, so that the mean of each variable in the training data becomes as close as possible to the mean of each variable in the test data. This is called a co-variate shift. These weights are determined with an optimization algorithm. The optimization problem is shown in formula 5.1 (Cortes et al., 2008). The data needs to be normalized before applying the KMM algorithm.

$$min_{\gamma}\ G(\gamma) = \left\| \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \gamma \Phi(x_i^{tr}) - \frac{1}{n_{te}} \sum_{i=1}^{n_{te}} \Phi(x_i^{te}) \right\| \qquad (5.1)$$

where γ is the vector with weights which is determined by the optimization algorithm, G(γ) is the distance between the means of the weighted training data and the testing data which is minimized, x_i^{tr} are the independent variables only of the training data, x_i^{te} are the independent variables only of the test data, n is the number of observations in the training or test data, and $\Phi(x)$ is the kernel function that maps x to a reproducing kernel Hilbert space (RKHS) (Berlinet & Thomas-Agnan, 2004). A weakness of KMM is that it gives equal importance to all independent variables. Another weakness of KMM is that the algorithm only matches the mean but not the entire distribution between training and test data. There are many different solutions to get to a matching mean. Some might not lead to a better match of the entire distribution, for example when large weights on error prone outliers are applied to shift the mean. Since the damage models are sensitive to extreme values, it would be desirable that the sample selection bias correction method leads to a better match of the entire distribution.

Cluster-Based Estimation (CBE)

In cluster-based estimation (CBE), the entire dataset (training and test data) is first split into several clusters. These clusters are made by combining the independent variables of the training and test data and then applying an unsupervised learning algorithm to find clusters of observations that lie relatively close together. After the clusters are identified, both the training and test data are split into these clusters. The weights are then determined in such a way that each weighted cluster appears as frequently in the training data as it appears in the test data. See formula 5.2:

$$CW_x = \frac{\frac{N_{x,test}}{N_{test}}}{\frac{N_{x,train}}{N_{train}}} \qquad (5.2)$$

Formula 5.2, where CW_x is the cluster weight to be applied on each sample in the training data that belongs to the specific cluster. $N_{x,test}$ is the number of samples in the test data that belong to that cluster, N_{test} is the total number of samples within the test data. $N_{x,train}$ and N_{train} are the same but for the training data.

The unsupervised learning algorithm k-means clustering is applied. This algorithm splits the data in K clusters based on the nearest means by placing K points in the spectrum of the data. It then clusters each data point based on which of the K points it is most close to (Kanungo et al., 2002). The algorithm then optimizes the position of the K points in such a way that the total distance of all data points to the locations of the K points is minimized. The data needs to be normalized before applying the algorithm.

Just as the KMM method, the disadvantage of CBE that all variables are equally important, while in fact the variables differ in their importance for predicting the damage. For example, wind speed is often a more important variable than for instance the economic growth of a municipality, in the case of wind damage estimation. Since all variables are assumed to have the same importance in the clustering, this may lead to clusters that are not particularly relevant for reducing the sample selection bias.

Single variable distribution matching (SVDM)

The CBE method is normally trying to match the distributions of all different variables. Some of these variables are however less important for the damage estimation than others. The CBE method is unaware of this difference in importance and will only try to match all available variables with equal importance. Matching the distributions for each variable perfectly is not possible on such small data sets, so compromises are made. These compromises reduce the quality of the match in the more important variables and therefore may reduce the model performance compared with a method that focusses on the most important variable.

Therefore, we introduce a special configuration of the Cluster Based Estimation (CBE), which we call single variable distribution matching (SVDM). This method makes use of the expert knowledge on the most important variable for the damage model. This works by just supplying the CBE method with the most important variable, such as water depth for floods or the wind speed for typhoons.

A disadvantage of adjusting for the most important variable only is that sometimes a transfer needs to be made over multiple variables. For example, a transfer in both building styles and water depth would be impossible with this approach. It is however possible to optimize this method by using several important variables rather than only the most important one. Such configurations are not explored in this research and rather only the two most extreme configurations are applied: i.e. all variables in common CBE or only the most important variable in the case of SVDM.

5.2.1.2 Synthetic data generation

The sample selection bias correction methods sometimes generate high weights for specific observations, for instance when one observational value is 30 times more important than another. Generating synthetic training data by resampling can create new data with similar statistical characteristics to the weighted training data. This results in many data points similar to the observation with the high weight rather than one specific point with a very high weight. This can be done with synthetic data generation techniques that are applied for example to meteorological or river discharges data (Diermanse et al., 2012).

This synthetic data generation approach to smooth out the high weights has been inspired by a similar method called Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002). This technique helps to correct imbalanced training data in classification problems, for instance, when rare observations in the training data need to be predicted.

Synthetic data is generated by building a statistical model that represents the sample selection bias corrected training data. From this statistical model new data points are sampled. This procedure can be summarized as follows:

- A Kendall's rank correlation matrix (T) is derived from the training data. The matrix is a square matrix with the size of the number of variables.
- A matrix P is derived through Cholesky decomposition, in which PxP-1 = sin(phi T/2) (Fang et al., 2002).
- For each variable, sample values with the standard normal distribution function are generated using its mean and standard deviation.

- Correlation is introduced between these individual samples. Such correlated samples are calculated based on multiplication between the transpose of matrix P and the sample values for each variable.
- To go from the normally distributed to the originally observed distributions in the training data an inverse transformation is applied to the normalized correlated sample based on the variable's empirical distribution.

5.2.1.3 Machine learning techniques

Machine learning (ML) is a field of artificial intelligence that provides computer systems the ability to learn from data without being explicitly programmed. ML algorithms are classified into i) supervised learning ii) unsupervised learning and iii) reinforcement learning.

This study focuses on the application of supervised learning algorithms (Praveena & Jaiganesh, 2017) to build models that can explain the complex relationships between damages and the variables that can explain damages, such as water depth or wind speed. We applied Random Forest and Artificial Neural Networks in this study. Random Forests are chosen because they have a good track record in damage modelling (e.g. Wagenaar et al., 2017; Schröter et al., 2018; Wagenaar et al., 2018; Sieg et al., 2017; Ganguly et al., 2019; Amadio et al., 2019), Artificial Neural Networks have also been used before in flood damage models (Ganguly et al., 2019; Amadio et al., 2019), and in this study they were selected because of their ability to extrapolate and at the same time find complex non-linear relationships. Table 5.1 provides a comparison between the ML methods. The K-means unsupervised learning algorithm is applied within the Cluster-Based Estimation sample selection bias correction technique.

Random Forest

Random Forest, a ML method developed by Breiman (2001), has been used in flood damage modelling (e.g. Wagenaar et al., 2017; Schröter et al., 2018; Wagenaar et al., 2018; Sieg et al., 2017; Ganguly et al., 2019; Amadio et al., 2019). Random Forests are ensembles of regression trees where the data for each tree is generated by a bootstrapping resampling method. In each tree, branches are formed by splitting the dataset based on binary recursive partitioning, for instance, a partition of data based on whether the average wind speed is higher than a certain value. The Random Forest algorithm does not use all explanatory variables for each tree, but it seeks the best splits within a limited number of sampled explanatory variables. The number of sampled features is the square root of the total number of features in the datasets. The best splits refer to regression trees that split the training data in such a way that there is as little variation as possible within the resulting leaves. The predicted value for the entire Random Forest is the mean value of the predictions from each regression tree.

A disadvantage of a Random Forest is that they can never make a prediction that is higher than the values seen in the training data, hence it cannot extrapolate (Tyralis et al., 2019). This is because each regression tree has a set number of leaves. When making a new prediction it will place the prediction in an existing leaf. It cannot create a new leaf with a higher damage value. In a damage model transfer setting this inability to extrapolate can be a disadvantage as extrapolation is sometimes required. An advantage of Random Forests is that they can make probabilistic predictions, which is however not utilized in this paper.

Artificial Neural Network

An Artificial Neural Network (ANN) is a ML framework inspired by how the human brain processes information (Hassoun, 1995). It was first introduced by Rumelhart et al. (1986), ANNs gain knowledge through learning the relationships between variables in a dataset without any given information about the system. The model built based on ANNs consists of several (hidden) layers of neuron-like processing units connected with each other. Each neuron is connected to all other neurons in the layer before it and after it. The connections work through coefficients that weigh each value that comes through the neuron. The coefficients of the neurons are determined with an optimization algorithm that minimizes the error on the training dataset. A strength of ANNs is that they can simulate complex non-linear patterns. Larger ANNs with more neurons can represent more complex non-linear patterns but are also more prone to match the training data so well that it works poorly on new cases (overfitting). The model built in this study is based on a multilayer perceptron (MLP) ANN, which consists of an input layer, two hidden layers, and an output layer (prediction). For transferring multi-variable damage models, ANNs may have an advantage over Random Forests in that they can extrapolate. In an ANN inputs are multiplied with coefficients. When the input value in the test data (e.g. water depth) are larger than the inputs in the training data the predicted value will be larger also. A general disadvantage of ANNs is that their predictions are deterministic and hence less suitable for applications that would benefit from having probabilistic estimates.

Table 5.1: Comparison of the Random Forest (RF) and Artificial Neural Networks (ANN) machine learning methods.

RF	ANN	REFERENCE
capture non-linear relationships		Nawar & Mouazen, 2017
overfitting may occur when too many splits in a tree are made	overfitting may occur when too many hidden layers are included	Ahmad et al., 2017a; Breiman, 2001
has few tuning parameters, which are often insensitive	has more tuning parameters	Ahmad et al., 2017a; Breiman, 2001
When applied to the same data set, typically, faster to train	When applied to the same data set, typically, slower to train	Ahmad et al., 2017b
Cannot extrapolate	Can theoretically extrapolate	Tyralis et al., 2019
Provides probabilistic predictions	Provides deterministic predictions	

5.2.2 Case studies

A case-study approach was used to quantitatively assess the improvement of the spatial and temporal transferability of damage models based on an ANN or a RF upon applying the three sample selection bias correction methods. Two case studies were used to allow a deeper insight into the application of damage models at two different spatial scales: macro level (municipalities) and micro level (buildings).

Macro level damage models predict the damage based on the aggregated data within one administrative boundary (e.g. village, district). This detail level is sufficient for many applications and the data is easier to collect. For the macro-level, a case study with typhoons in the Philippines on municipality level was adopted. The models in this paper are an extension of macro-level data driven multi-variable models that were developed to support humanitarian aid organizations with the prioritization for distributing humanitarian aid after or just before a typhoon. The models aim to provide timely and unbiased information after a disaster, which are often difficult to obtain using current practices (field surveys).

Micro level damage models, on the other hand, predict the damage on disaggregated level (e.g. per building). Micro level models are often used for disasters that require a detailed spatial resolution such as in our case for damage from fluvial floods in Europe. Such level of detail is required in insurance applications when risk premiums need to be determined per building, or for flood mitigation policies when measures on building level are considered. Even though for many such applications the results are later aggregated, the calculations are often done on micro level because macro models can lead to considerable spatial uncertainty (Wagenaar et al., 2016).

The data used has been selected after an assessment of the data quality on different attributes, i.e. timeliness, source (reliability), accuracy and granularity/spatial coverage (van den Homberg et al., 2018) as will be explained for each case study. Obviously, the data for both the independent and dependent variables needs to be available at the same granularity and spatial coverage. Table 5.2 summarizes the characteristics of the two case studies. In both cases the data is always used in a transfer setting. That means the data is applied on an event or a location that wasn't part of the training data.

Apart from the spatial scale, the cases use different types of variables, damage mechanisms and type of transfer. The macro case study has more vulnerability type variables such as poverty and building materials, and has in some cases more damage mechanisms, such as floods due to a storm surge caused by the typhoon. The transfer for the macro model was over time since all data comes from the same larger study area. In the micro model there is both a time (different events) and space transfer (between The Netherlands and Germany). These large differences are an ideal test to see whether the sample selection bias correction techniques work under different circumstances.

Table 5.2 Characterization of case studies

Case study	*Level*	*Event*	*Dataset*	*Applications*	*Data preparation and Machine learning*
Typhoons, The Philippines	Macro, Relative damage at municipality level	12 typhoons 2012 - 2016	1600 records of percentage totally destroyed buildings in municipalities	Area prioritization on distribution of humanitarian aid	Sample correction bias methods Synthetic data generation (3000 to 10000 samples) cross validation between events based on ANN/RF
Fluvial floods, The Netherlands and Germany	Micro, Relative (content and structural) damage at building level.	1 flood event 1993 6 flood events 2002 - 2013	The Netherlands: 4398 monetary residential damages Germany: 895 monetary residential damages	Insurance, risk reduction and mitigation, cost-benefit analysis	Sample correction bias methods Synthetic data generation (3000 to 10000 samples) Transferring Dutch model to Germany based on ANN/RF

5.2.2.1 Macro level model: Philippines typhoons

On average around twenty typhoons strike the Philippines annually and more than half of them make landfall in the country (reliefweb, 2018). Typhoon Haiyan (local name Yolanda), which hit the Philippines in 2013, is considered one of the strongest tropical cyclones ever recorded. The fatalities caused by the typhoon amounted to about 6,000 people, around 14 million people were affected and more than 1 million houses were damaged (World Bank, 2014).

510, an initiative of the Netherlands Red Cross collated the typhoon damage data in this case study through desk research and in-country visits of key stakeholders. The purpose of collating this data is to populate 510's community risk assessment dashboard and to develop a model that can be used to predict the areas with the highest damage either just before the disaster to trigger early action. or just after the disaster to improve efficiency in the aid distribution process.

Data

Data has been gathered on twelve typhoons in the Philippines at the municipality level. The median number of households in a municipality is around 6,600. The dataset contains about 1,600 damage records, with 40% of those damage records corresponding to the two typhoons that cover the largest extent. This does not necessarily mean that they have the largest aggregate damages.

The vulnerability and exposure variables in a municipality are the same for all typhoons while the hazard features are specific to a typhoon. The vulnerability and exposure may have changed over time in the period from 2012 to 2016, due to e.g. population growth and land use change. These changes, however, are typically relatively slow. Recovery efforts are an exception because damages could be lower in an area that was recently affected and hasn't recovered yet. This can be a source of variation in the data but is expected to be limited.

The dataset collected by the Red Cross consists of more than 40 variables from which damage is to be predicted. Table 5.3 presents the variables that were used to build the damage models for the macro case study. It is essential to have data on these independent variables with national spatial coverage and at the same administrative levels. The municipality level was chosen as the smallest geographic level because this is the lowest resolution on which all the data is available.

Table 5.3: Variables available for the macro case study.

VARIABLE NAME	UNIT	SOURCE(S)	REMARKS (MODEL SCALE)
COMPLETELY DESTROYED BUILDINGS (DAMAGE)	%	National DRR and Management Council (NDRRCM)	Percentage of the houses that are entirely destroyed and unfit for habitation or without any remaining structural features. Data collected for Emergency Shelter Assistance program (DSWD, 2019)
AVERAGE WIND SPEED	mph	Tropical Storm Risk (UCL, 2018)	Maximum 3 seconds sustained gust speed over the event in the particular municipality. Every municipality has a unique wind speed calculated based on the forecasted maximum wind speed on the track and the method from Holland (1980) to calculate it for the specific municipality.

ACCUMULATED RAINFALL	mm	Meteorological data from Global Precipitation Measurement (GPM) (Huffman et al., 2015)	Total accumulated rainfall during the typhoons period from satellite data (Huffman et al., 2015).
NUMBER OF HOUSEHOLDS		Philippines National Census	2010 data, unique value available for each municipality in the country.
POPULATION DENSITY	people/ km2	Philippines National Census	2010 data unique value available for each municipality in the country.
AREA	km2	GIS analysis	Area within the official municipality boundaries.
ELEVATION (AVERAGE AND WEIGHTED ON POPULATION)	m	STRM (NASA, 2013)	30-Meter Elevation Data
SLOPE	m/m	SRTM (NASA, 2013)	Based on QGIS slope module applied to 30-Meter Elevation Data. (QGIS slope module, 2020)
ROOF TYPES (WOOD, IRON, STRAW, CONCRETE, SEMI-CONCRETE) IN AN AREA	%	Philippines National Census	Based on 2008 data, unique value available for each municipality in the country.
WALL TYPES (CONCRETE, MAKESHIFT, WOOD, CONCRETE, IRON, BAMBOO) IN AN AREA	%	Philippines National Census	Based on 2008 data, unique value available for each municipality in the country.
POPULATION UNDER POVERTY LINE	%	Philippines Statistics Authority	Available per province, each municipality has the province value.
LENGTH COASTLINE	m	GIS analysis	Based on official municipality boundaries.
RUGGEDNESS	m	SRTM (NASA, 2013)	This is the Terrain Ruggedness Index, defined as the mean difference between a central pixel and its

			surrounding cells, calculated on 30m SRTM elevation data with the QGIS TRI module (QGIS TRI module, 2020)
POPULATION LIVING 500, 1000 AND 1500M FROM THE COAST	%	WorldPop (Gaughan et al., 2013)	Based on a GIS analysis combined with WorldPop data (worldpop.org.uk)
ECONOMIC GROWTH	%	Philippines Statistics Authority	Annual growth for the year 2018. Available per province, each municipality has the province value.
POPULATION GROWTH	%	Philippines Statistics Authority	Annual growth for the year 2018. Available per province, each municipality has the province value.

Model setup and validation

The paper proposes to build data-driven damage models that can be part of a model set-up used for operational purposes on newly predicted events. The evaluation can be carried out by using one of the observed typhoons as test data and use the rest as training data. The damage caused by a historical typhoon is predicted by a model built based on the data from the other eleven typhoons. As there is data about twelve typhoons recorded in the dataset, twelve prediction models were built in total with each typhoon serving once as the test data for which the model is then tested.

Data-driven damage models were developed to predict the percentage of completely destroyed buildings in an affected municipality based on the variables shown in Table 3. The most interesting aspect about the damage data is that the average of damage varies between the twelve typhoons. The average value over all typhoons is 6% of the buildings completely destroyed, which is nearly 6 times smaller than the average for typhoon Haiyan.

From Figure 5.2, it can be seen that the distribution of the damage to buildings caused by typhoon Haiyan is much higher than for the other 11 typhoons. This indicates that the damage data from the other 11 typhoons that is used to build the prediction model for Haiyan are not fully representative for this typhoon and hence a major model transfer is required that includes extrapolation. This is a typical example where advances in the transferability of models may improve damage predictions.

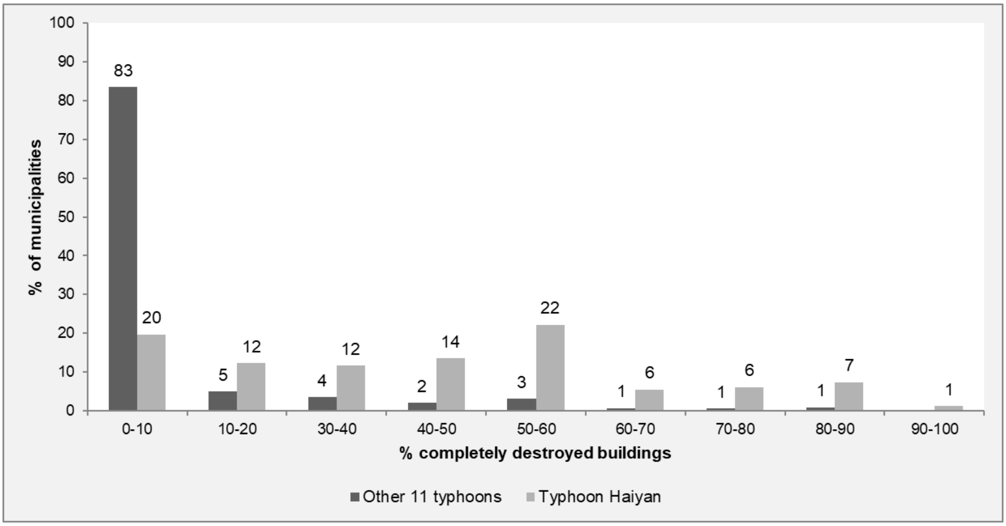


Figure 5.2: The distribution of completely destroyed houses per municipality for the Haiyan typhoon compared to the other typhoons that were used to build a model for Haiyan.

5.2.2.2 Micro level model: European flood damage models

Damage data and independent variables for the micro level case study were selected for six past river flood events in Germany between 2002 and 2013 and for one river flood event in the Netherlands in 1993. This data has been used for several data-driven models in the past (Wagenaar et al. 2017, Wagenaar et al., 2018, Schröter et al., 2014; 2018, Merz et al., 2013. In the current micro level case study, a multi-variable flood damage model made based on Dutch data is transferred to Germany. The same model transfer was done in the paper by Wagenaar et al. (2018), which showed that this model transfer could potentially be improved, as it was the model with the lowest performance, owing to the low variability of the damage data in the 1993 flood event in The Netherlands. The expectation therefore is that the model can be improved by correcting for the known sample selection bias. The flood damage model predicts the relative damage on building level based on the variables shown in table 5.4.

Data

The Dutch training data in this case study is derived from observed flood damages after the 1993 flood in the Meuse River in Limburg reported in WL Delft (1994), supplemented with data on building and flood characteristics documented in Wagenaar et al. (2017).

Table 5.4: Variables used in the micro level case study (more information in Wagenaar et al., 2018

VARIABLE NAME	UNIT	SOURCE DUTCH DATASET	SOURCE GERMAN DATASET	REMARK
RELATIVE BUILDING DAMAGE	-	Inspection and building value estimate	Phone interview	Relative to potential damage.
RELATIVE CONTENT DAMAGE	-	Inspection and content value estimate	Phone interview	Relative to potential damage.
WATER DEPTH RELATIVE TO GROUND FLOOR	m	Inspection	Phone interview	
BUILDING TYPE		Inspection	Phone interview	2 types available, attached or unattached.
FOOTPRINT AREA BUILDING	m2	Cadastre	Phone interview	
WATER DEPTH RELATIVE TO DEM	m	Model		For German data equal to water depth relative to floor.
BASEMENT		Inspection	Phone interview	
HOUSEHOLD SIZE	#	Inspection	Phone interview	
FLOW VELOCITY	m/s	Model	Phone interview	For German data estimated from score
BUILDING AGE	Year	Cadastre	Phone interview	
FLOOR AREA FOR LIVING	m2	Cadastre	Phone interview	
FLOOD DURATION	hour	Hydro-dynamic Model	Phone interview	
RETURN PERIOD	year	Statistical model	Statistical model	Definition in: Wagenaar et al. (2018)

The model is applied to predict the damage from six different flood events in Germany. Damage from these floods including relevant building and flood characteristics were collected using phone interviews (Thieken et al. 2007, Kreibich et al. 2017). The German dataset contains a wide range of values for the different flood characteristics and circumstances (Kreibich et al. 2011, Kienzler et al. 2015), the Dutch data is on the other hand more homogenous because they are based on only one flood event (Wagenaar et al., 2018).

Model setup and validation

There are 895 damage observations from the German data that can be used to test the model made based on the 4398 damage observations from the Dutch data. To reduce the randomness in the predictions due to the specific selection of training samples, bootstrapping is applied (Efron and Tibshirani, 1993). In bootstrapping, a random sample of the training data is taken to train the model, and then a random sample of the test data is taken to test the model. Samples are taken with replacement. This is repeated several times, so that many models are trained and tested on such subsets of the data. For each bootstrap run, 4000 training samples from the Dutch data and 350 testing samples from the German data were randomly picked. Bootstrapping reduces the chance that a difference between the two samples is due to randomness rather than because of an improvement in the prediction method. For the RF, 100 bootstrap samples were taken. On the other hand, only 20 bootstrap samples were taken for ANN due to the greater calculation time. Less samples were taken for the ANNs, as differences between the calculated errors were shown to be minor, while the calculation time was much longer for the ANN than for RF.

5.2.3 Model parameters

Damage models built based on Random Forest and Artificial Neural Networks have been developed using the Python 2.7 library “Sci-Kit learn” (Pedregosa et al., 2011). For the damage model based on Random Forests, a hundred regression trees were grown. More regression trees need more computation time but also typically give better results. This improvement from adding more trees becomes negligible after a certain number of trees. For this study the same number of trees are applied as in Wagenaar et al. (2018) and the model errors couldn’t be reduced by adding more than a hundred trees. The number of splits and minimum number of observations per leaf were optimized. For the prediction model based on ANN, learning rates and number of neurons in the first hidden layer were optimized. The number of neurons in the second hidden layer was fixed to be half of the neurons in the first layer.

This optimization was carried out by randomly splitting the dataset into 60:40 for the training and test set. The tuning of the parameters for both models was carried out to result in the smallest MAE on validation data that didn't involve a model transfer (splitting the training data randomly).

The CBE and SVDM methods have one parameter to tune: the number of clusters used. This was chosen to be twelve clusters for both case studies. The Kernel Mean Matching method has only one parameter to be optimized also: the kernel to be used. Linear kernel was chosen because of its simplicity. For SVDM, the most important variable to predict the damage chosen was wind speed for the macro model and water depth for the micro model. Both variables are widely used in single variable damage models (e.g. Pielke, 2007; Merz et al., 2010; Gerl et al., 2016). Furthermore, the feature importance analysis carried out within the random forest confirms this choice.

For the synthetic data generation, the number of synthetic data points to be generated can be optimized. More synthetic data points generated generally gives better results, but after a specific point they don't considerably affect the results anymore. For the macro model, the number of synthetic data points to be generated is always twice the weight of the training set after the sample selection bias correction methods are applied. This is based on a minimum weight of one, so the sample selection bias correction increases the number of data points. This typically turns out to be between 3000-10,000 synthetic data points. For the micro model, a simplified approach was applied with a fixed number of samples because the training set is always the same size, this fixed number of samples is 5000. The number of samples to be taken was estimated based on increasing the number of samples until the evaluation metrics would no longer improve.

5.2.4 Evaluation metrics

To evaluate the model performance three different error metrics were used: Mean Absolute Error (MAE), Mean Bias Error (MBE) and Symmetric Mean Absolute Percentage Error (SMAPE). Table 5.5 shows the formulas for the different evaluation criteria.

The MAE is suitable to evaluate the accuracy for individual predictions. This is important when the individual model results need to be applied, for example for insurance or for macro level models. The MBE shows whether there is a bias in the model, for instance whether it consistently makes over or underestimations. This is important when the aggregated results are used. For example, in a micro model used for a cost benefit analysis for an infrastructure investment, only the total sum of all predictions is important rather than individual prediction per building. In such a case, the MBE is the most important evaluation criteria. The SMAPE is used in the same manner as the MAE but is a relative error metric. This allows to compare the errors of different order of magnitude events. For example, some models have predicted damages in the order of 50-80% while others have a maximum of 20%. A 20 percentage point error on a damage of 80% is much lower relatively, than a 20 percentage point error on a damage of 10%.

For the macro model, the errors were evaluated for twelve different typhoons. Then the weighted mean of their errors was calculated. The weights were assigned based on the number of predicted damages in each model. For the MBE the absolute values are taken before the mean is calculated over the 12 events. This is done in order to ensure that a positive bias in one test cannot cancel out a negative bias in another test. Consequently, all bias errors are positive. In order to obtain the mean that represents the quality of the twelve models, the criteria to evaluate errors should be independent from the extent of the damage they predicted. SMAPE is particularly useful for this case study, as the errors for different models are compared with each other. For the micro case the variation between the damage cases is less extreme and therefore a SMAPE approach is not necessary.

In this paper no evaluation metrics are applied to validate the quality of the probabilistic estimates of the RF and to see whether these probabilistic estimates improve because of sample selection bias correction methods. This is not done because ANNs aren't able to make such probabilistic predictions. This could be a topic for future research.

Table 5.5: Criteria applied to evaluate the model performance. RL_{sim} is the simulated/predicted relative loss. RL_{obs} is the observed relative loss.

EVALUATION CRITERIA	FORMULA
MEAN ABSOLUTE ERROR (MAE)	$MAE = \frac{1}{N}\sum \lvert RL_{sim,n} - RL_{obs,n}\rvert$
MEAN BIAS ERROR (MBE)	$MBE = \frac{1}{N}\sum RL_{sim,n} - RL_{obs,n}$
SYMMETRIC MEAN ABSOLUTE PERCENTAGE ERROR (SMAPE)	$SMAPE = \frac{\sum \lvert RL_{sim,n} - RL_{obs,n}\rvert}{\sum \lvert RL_{sim,n}\rvert + \lvert RL_{obs,n}\rvert}$

5.3. Results and discussion

Table 5.6 compares the performance of the predictions of the different ML models as measured by the evaluation metrics described in 5.2.4. It is apparent from the highlighted numbers in this table that the best performing models in both case studies and for all evaluation criteria always have some form of sample selection bias correction included. Furthermore, on the MBE evaluation criteria all sample selection bias correction methods always outperform the reference models. The improvements on the Mean Bias Error metric can be as large as 85% (e.g. MBE content damage), where many different sample selection bias correction methods result in large improvements. It is promising that the sample selection bias correction methods lead to improvements in both case studies, despite the large differences between the phenomena and data in the case studies, as discussed in section 5.2.2.3.

For the MAE metric, the results are a bit more varied. For the micro model, the improvements are minor. On the other hand, every sample selection bias correction method provides improvements for the macro model on the MAE criteria. The improvements for the SMAPE are however much smaller and are more in line with the improvements seen on the MAE for the micro model. Some sample selection bias correction methods are also not better than the reference models without sample selection bias correction for the SMAPE. The performance on the MAE for the macro model is mostly based on the model performance on the extreme observations, because these observations have large errors, improving them has a relatively large impact on the MAE. For the SMAPE error metric the large and small damage observations have a more equal weight in the error metric calculation. The sample selection bias correction methods therefore seem to be most relevant to predict outlier observations. These results seem to be consistent with the general idea that the sample selection bias correction is mostly suitable for extreme observations, which is very relevant for some of the applications of damage models.

Table 5.6: Performance of different models for both the micro and macro case studies. The best performing model setup is made grey and bold. The 2nd and 3th best performing methods are made grey.

Methods			Macro model			Micro model			
			% damaged buildings			Building damage		Content damage	
Machine Learning Methods	Sample Selection Bias Correction	Synthetic Data Generation	MAE	MBE	SMAPE	MAE	MBE	MAE	MBE
RF			5,53	3,86	0,672	0,100	0,089	0,213	0,207
ANN			6,29	4,15	0,668	0,104	0,097	0,218	0,212
RF	KMM		5,27	3,48	0,618	0,099	0,082	0,211	0,202
ANN	KMM		5,78	3,25	0,674	0,109	**0,014**	0,204	**0,081**
RF	CBE		5,06	3,31	0,613	0,098	0,079	0,209	0,198
ANN	CBE		5,16	**2,53**	0,664	0,096	0,064	0,198	0,169
RF	SVDM		5,25	3,45	0,617	0,099	0,034	0,196	0,111
ANN	SVDM		5,69	3,53	0,661	0,097	0,085	0,211	0,205
RF	KMM	SD	4,69	3,29	**0,608**	0,095	0,065	0,198	0,171
ANN	KMM	SD	5,89	2,73	0,674	0,112	0,030	0,210	0,099
RF	CBE	SD	4,90	3,53	0,627	0,098	0,079	0,206	0,192
ANN	CBE	SD	5,63	3,40	0,672	0,095	0,055	0,195	0,153
RF	SVDM	SD	**4,43**	3,29	0,613	**0,095**	0,076	0,206	0,198
ANN	SVDM	SD	5,71	3,69	0,667	0,101	0,038	**0,194**	0,100

In theory these techniques should not work in a situation without a model transfer because there shouldn't be any bias in the data when the training and test data come from the same source (i.e. same variable distributions). The weights calculated by the sample selection bias correction methods should in that case be close to one and therefore the methods do not correct for anything. To test this, the best performing sample selection bias correction methods were also applied to settings without a model transfer. For the micro model, independent test data comes from the same source as the training data (Dutch data). For the macro model, all observations are put together and then split into training and test data. In this setting, the sample selection bias correction methods had hardly any influence on the results for the macro model (data not shown). A reduction was seen only in the MBE on the micro model, but without a model transfer this MBE is negligible (close to zero). Therefore, the reduction is very minimal in absolute terms.

The sample selection bias correction methods lead to a larger reduction in the MBE in combination with the ANN methods then in combination with the RF methods. Without sample selection bias correction methods, the ANN model performs less well than the RF model. This occurs consistently in both the micro and the macro case. The reason for this is not entirely clear, but we speculate that this could be due to the sensitivity of the different ML methods to the data.

5.3.1 Macro case study

For the Philippines case study, sample selection bias correction methods have considerably improved predictions from the twelve damage models. Figure 5.3 (left) visualizes an example of the improvement in predictions for the extreme typhoon Haiyan after employing the SVDM method in combination with synthetic data generation to the ANN. It shows that without sample selection bias correction the model consistently underestimates the damage as all estimates are below 30%. After implementing the sample selection bias correction method this consistent underestimation is largely solved and damages are predicted up to 60%, as they were also observed for typhoon Haiyan.

Figure 5.3 (right) provides an insight on how the different machine learning methods result in varying improvements. It can be seen from the figure that the ANN model results in more accurate predictions for the Haiyan typhoon compared to the RF model after the sample selection bias correction methods are applied. The results also further support the theory that a model built with an ANN is better able to predict the damage by extrapolation, compared to the RF model.

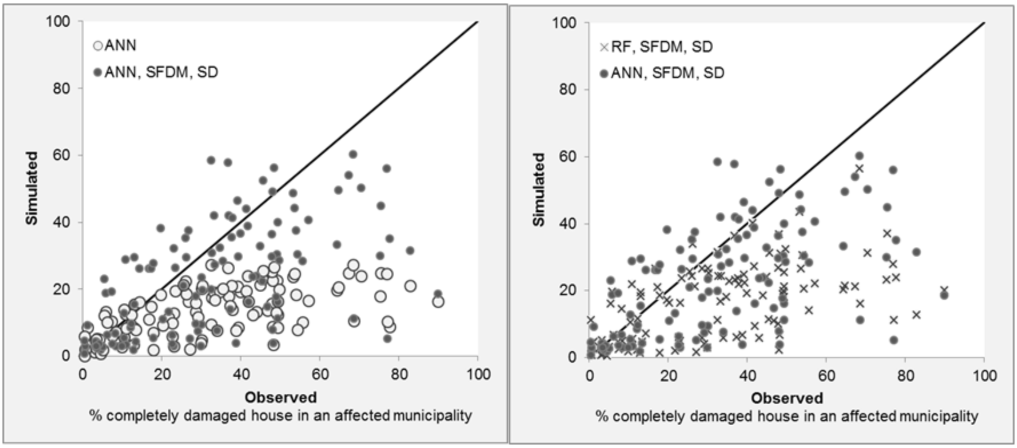


Figure 5.3: The performance per municipality for the model to predict damages for the Haiyan typhoon. Left) A comparison of the ANN method with and without sample selection bias correction and synthetic data generation. Right) A comparison of the RF and ANN methods with sample selection bias correction and synthetic data generation.

Table 5.6 shows that the predictions from twelve models built using RF as basis ML method provides the smallest errors on average. This implies that most of damage caused by other typhoons than an extreme typhoon such as Haiyan can be better predicted by a RF that can only interpolate and not extrapolate. This makes sense because the extrapolating capacity of ANN is not required for most of the data points, apart from data points of extreme typhoons.

A possible explanation for why the ANN models perform worse for average model results than the RF models is that RF works better on a relatively small datasets. Another likely explanation is that the ANN model is quite sensitive for parameter tuning while the RF model is not. The procedure for tuning the parameters could be improved. The tuning should not be carried out for all models at once based on the randomly split data (See section 5.3.3), but for each of the twelve models separately. The tuning of parameters that result in the smallest weighted mean error for the twelve models together then should be applied to all the twelve damage prediction models to be evaluated.

In general, the macro case study is limited by the lack of information on exposure and vulnerability variables. Adding more variables could be helpful. Also, the data on the explanatory variables was the same for all events regardless of the year in which the typhoon hit. Over time these characteristics may have undergone change, requiring changes in the variables. For example, houses might have been built back better after a typhoon with different materials. Especially locally this is expected to lead to some error for instance when large damages have occurred recently and people have responded by abandonment or much stronger building construction. These errors are however expected to have a negligible effect on the aggregated results of this case study.

5.3.2 Micro case study

Sample selection bias correction methods have reduced the MBE for all cases in the micro model case study. In four out of twelve cases this reduction is even larger than 50%. The MBE is the most relevant metric when the aggregated results of micro models are used. The MAE improvements for the micro model are rather small but in line with the SMAPE improvements of the macro model. This is probably because outliers have a smaller influence on the aggregated MAE metric for the micro model than for the macro model in which large differences between damages were presented. Another possible explanation is the difference in data quality of the micro model. The macro model consists of municipality averages while the micro model has values per building. The average values per municipality can correct overestimations and underestimations and hence the aleatory uncertainty is reduced. For individual building values, however, aleatory uncertainty is very high, and no such evening out of errors by averaging exists. This aleatory uncertainty cannot be reduced by sample selection bias correction methods and therefore the reductions in MAE are smaller in the micro model.

5.3.3 Performance of new sample selection bias correction methods

In this paper two innovations in sample selection bias corrections were introduced: using a single variable correction in the CBE method (SVDM method) and synthetic data generation. These innovations were compared to two other correction methods (KMM and CBE) with and without synthetics data generation.

Single Variable Distribution Matching (SVDM) method

The CBE method applied to only a single variable (SVDM) often performs better than the CBE method applied to multiple variables, according to the MAE criteria. The likely reason is that a better match can be made for the most important variable when variables of minor importance to the damage prediction are not considered for determining the weight. Including all variables in the CBE method leads to the best performance on the MBE criterium, compared to SVDM.

In practice a transfer will often need to be made over several important variables. For future research, multiple variables could be used to determine the weights for the training data. In this way a balance needs to be created between not diluting the influence of the most important variables on the weight, and correcting for biases in multiple variables rather than one. In addition, the user needs to determine whether absolute or average errors are most important for the application of the model.

Synthetic data generation method

The synthetic data generation combined with a sample selection bias correction method generally performs better than just the sample selection bias correction. This is especially the case for the MAE evaluation criteria. The reason this works is probably because ML methods can create very sharp decision boundaries. This means that when a few data points have very large weights the ML models can infer that only under the specific conditions of these data points the related high damage occurs but not with similar values. For example, a large damage could according the model only occur at 4m water depth but not at 3.9 or 4.1m. This is a form of overfitting. The synthetic data generation methods introduce some variation in these high weighted samples and hence increase the decision region for which the ML method assigns a high damage. This is the same reason why the similar SMOTE method performs well (Chawla et al., 2002).

The disadvantage of the synthetic data generation methods is that information inside the data might be lost while building the statistical model to draw synthetic data points from. A future method would be desirable that also increases the decision region but minimizes the loss of information from the original data. A possible approach that could be considered is the use of differential privacy techniques (Khatri, 2017). These techniques add small perturbations to the data to reduce privacy concerns. Recently, Khatri (2017) found that these perturbations work to prevent overfitting also.

5.4. Conclusions

Recent advances in damage models include data-driven methods to estimate damages caused by natural hazards. An important quality of such methods is their ability to capture complex, non-linear relationships between multiple variables related to hazard, exposure and vulnerability. However, data-driven method are usually limited by the availability and quality of the data required to build such models. As a result, transfer of the models (i.e. using data from one location to build a model for another location) is often required. This raises a problem, the sample selection bias, as the collected data is often not fully representative for the situation it needs to be applied on.

This study was undertaken to improve such methods to correct for this sample selection bias, and to evaluate the quality of the predictions. Such corrections were applied on two different case studies: (i) a macro level damage model for typhoons in the Philippines and (ii) a micro level damage model for European river flood damages.

Two machine learning (ML) techniques were used: Random Forests (RF) and Artificial Neural Networks (ANN). They were then improved by using the three different methods to correct the sample selection bias: Kernel Mean Matching (KMM), Cluster-Based Estimation (CBE), Single Variable Distribution Matching (SVDM), which apply weights to the training data. As sometimes very high weights are assigned to specific observations, additionally, a statistical model was built to generate a larger set of synthetic training data before the ML techniques were applied.

We conclude that multi-variable data-driven damage models should correct for the sample selection bias that arises from a model transfer setting, as especially on the mean bias error (MBE) large reductions are possible, amount to more than 30% error reduction. For a large model transfer (e.g. data from small typhoon to predict damages from an extreme typhoon), the ANN method seems to further improve the predictions compared to the RF method, probably because the method is better capable of extrapolation. These sample selection bias correction methods are especially important in reducing Mean Bias Errors for the micro models and lead to up to 50% reduction on MBE, compared to reductions up to 10% on the MAE. For macro models the correction methods are shown to also reduce the Mean Absolute Errors, with a reduction up to 20%.

Synthetic data points generated from the sample selection bias correction methods are shown to considerably improve the models for the Mean Absolute Error criteria, and more than half of the improvement is introduced by the synthetic data for the MAE metric. Future studies that correct for a sample selection bias should therefore consider extending the dataset using synthetic data generation after the sample selection bias correction.

This study shows that in the future, data-driven damage models should consider sample selection bias correction methods when a model transfer is required. This helps to reduce the MBE and to better predict outlier observations. To correctly predict these outlier cases synthetic data generation or similar techniques can be used. In transfer cases where the simulation of extreme values beyond the observational data is required, machine learning techniques should be considered that can allow extrapolation, such as ANN in this study.

Further research could help establish a reliable impact-based forecasting system based on data driven multi-variable models. This system would be of great help for several sectors, ranging from insurance industry to humanitarian aid organizations. The insurance industry can apply this model to estimate risk premiums. Humanitarian organizations can use data-driven predictions to prioritize faster and better their preparation and aid distribution process in the early warning /early action phase, and after a disaster strikes.

Acknowledgements

We would like to thank Ferdinand Diermanse (Deltares) for his help with the synthetic data generation from statistical models. Also, we like to thank Aklilu Teklesadik (510) and Jannis Visser (510) for their role in the data collating for the case study in the Philippines. Furthermore, we would like to thank Jonathan Nuttall (Deltares), Hans Korving (Deltares) and Danish Shrestha for their help with the Machine Learning methods.

This research was supported by the Dutch ministry of economic affairs and climate for their funding through the Deltares strategic research programs: Flood Risk Strategies, Enabling Technologies and Consequences of Extreme Weather.

This research was further supported by the European Union's Horizon 2020 research and innovation programme, through the IMPREX project (Grant Agreement number 641811). The surveys to collect the empirical damage data in Germany were supported by the German Research Network Natural Disasters (German Ministry of Education and Research (BMBF), no. 01SFR9969/5), the MEDIS project (BMBF; no. 0330688) the project "Hochwasser 2013" (BMBF; 13N13017) and by a joint venture between the German Research Centre for Geosciences GFZ, the University of Potsdam and the Deutsche Rückversicherung AG, Düsseldorf. This research also received support from the NWO-VICI grant number 453-13-006.

6. HOW MACHINE LEARNING WILL CHANGE FLOOD RISK AND IMPACT ASSESSMENTS

This chapter is based on the paper "Invited perspectives: "How machine learning will change flood risk and impact assessment" written together with Alex Curran (Delft University of Technology), Mariano Balbi (University of Buenos Aires), Alok Bhardwaj (Nanyang University of Technology), Robert Soden (Columbia University/ World Bank), Emir Hartato (Planet), Gizem Mestav Sarica (Nanyang University of Technology), Laddaporn Ruangpan (IHE Delft), Giuseppe Molinario (World Bank) and David Lallemant (Nanyang University of Technology). It is published in the journal Natural Hazards and Earth Science Systems (NHESS). The reference is: Wagenaar, D., Curran, A., Balbi, M., Bhardwaj, A., Soden, R., Hartato, E., Mestav Sarica, G., Ruangpan, L., Molinario, G., and Lallemant, D.: Invited perspectives: How machine learning will change flood risk and impact assessment, Nat. Hazards Earth Syst. Sci., https://doi.org/10.5194/nhess-2019-341.

Abstract

Increasing amounts of data, together with more computing power and better machine learning algorithms to analyse the data are causing changes in almost every aspect of our lives. This trend is expected to continue as more data keeps becoming available, computing power keeps improving and machine learning algorithms keep improving as well. Flood risk and impact assessments are also being influenced by this trend, particularly in areas such as the development of mitigation measures, emergency response preparation, and flood recovery planning. Machine learning methods have the potential to improve accuracy as well as reduce calculating time and model development cost. It is expected that in the future more applications become feasible and many process models and traditional observation methods will be replaced by machine learning. Examples of this include the use of machine learning on remote sensing data to estimate exposure or on social media data to improve flood response. Some improvements may require new data collection efforts, such as for the modelling of flood damages or defence failures. In other components, machine learning may not always be suitable or should be applied complementary to process models, for example in hydrodynamic applications. Overall, machine learning is likely to drastically improve future flood risk and impact assessments, but issues such as applicability, bias and ethics must be considered carefully to avoid misuse. This paper presents some of the current developments on

the application of machine learning in this field and highlights some key needs and challenges.

6.1. Introduction

Exponentially increasing computing power and data, as well as rapidly improving machine learning algorithms to analyse this data have been changing many aspects of our lives (Manyika et al., 2011). These trends are expected to continue and will undoubtedly keep affecting many scientific, commercial and social sectors (Manyika et al., 2011). Flood risk and impact assessments are no exception to this trend. Flooding yearly affects more people than any other natural hazard types (Jonkman, 2005) and the impact and frequency of flooding events is expected to increase in the future due to urban development and climate change (Kundzewicz et al., 2014). It is therefore an opportunity for researchers and flood managers to tap into the potential of machine learning, taking advantage of their strengths while being cognisant of their limitations. It is also important to anticipate improvements in the capabilities of machine learning methods, so as to plan for forthcoming changes in flood modelling.

When assessing the interaction between floods and society, three different components can be recognized: exposure, hazard, and impact (Kron, 2002). Exposure refers to the characteristics of the people and assets that can be affected by flooding. Hazards are the physical characteristics of a flood such as the extent, water depth, duration and flow velocity. Impacts are the effects the hazard has on the exposure. To assess these three components, we make the distinction between flood risk, as the probabilistic analysis of the potential (predictive) impacts of floods and flood impact assessment, as the post-event assessment of (descriptive) impact from an actual flood event. Table 1 provides examples of predictive and descriptive assessments in relation to the hazard, exposure and impact components. The scope of this paper is limited to the predictive and descriptive assessments shown in table 6.1 and doesn't include potential uses of machine learning in risk awareness or communication strategies.

Table 6.1: Overview of different types of flood risk and impact assessments

	PREDICTIVE	DESCRIPTIVE
EXPOSURE	Urban growth modelling	Identification of current built-up area
HAZARD	Flood modelling	Mapping current and past floods
IMPACT/VULNERABILITY	Forecasting impact (e.g. damage)	Assessing flood impacts (e.g. damage) after they have occurred

Flood risk and impact assessments have many different applications. A useful paradigm to look at these different applications is the 'disaster management cycle' (Khan et al., 2008; National Research Council, 2006) (figure 6.1). This cycle delineates the phases between events, i.e. the immediate response to an event, the long-term recovery, the mitigation to prevent future events and the preparation prior to a new forecasted event.

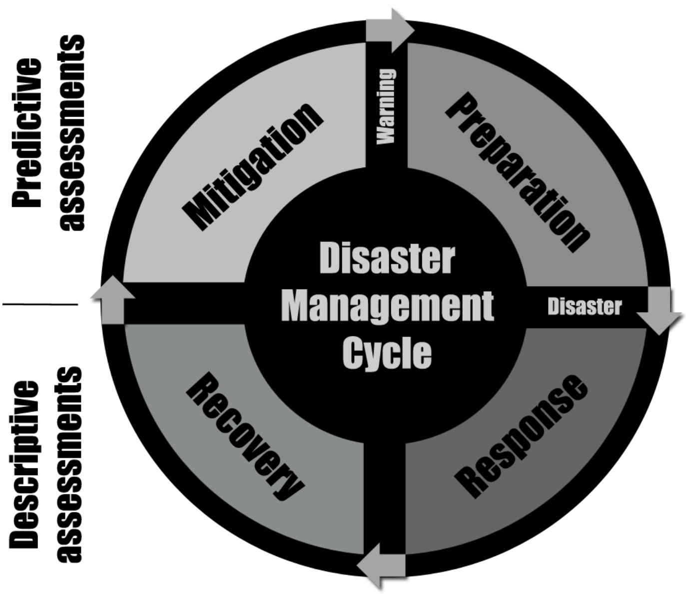


Figure 6.1: Disaster management cycle, a common paradigm tool.

In the response phase, the focus is typically on descriptive hazard, exposure and impact assessments (e.g. Klemas, 2015), sometimes complemented with predictive models if the event descriptive information isn't available yet (e.g. a predictive model estimating the number of people affected can be fed by a descriptive hazard model of the flood extent). The challenge in this phase is mostly data reliability. In the recovery phase, descriptive assessments are often used for payouts (e.g. indemnity insurance), and one of the main challenges is ensuring these payouts are timely and reliable. In the mitigation phase, probabilistic predictive models are used (e.g. Wagenaar et al., 2019), typically for the design of risk-reduction interventions ranging from protective infrastructure to insurance products. The challenge in this phase is model reliability and uncertainties about future developments (e.g. uncertainty in future exposure). In the preparation phase, predictive models are used for emergency planning (e.g. Coughlan de Perez et al., 2016), where the challenge is the reliability, availability and communication of data. Machine learning is capable of generating more reliable and faster models that can help solve some of the current challenges in the disaster management cycle but could also provide new opportunities (GFDRR, 2018).

Machine learning algorithms can find patterns in data and use these patterns to make predictions about new data (Bishop, 2006). For example, when providing a machine learning algorithm with aerial images of either urban or rural areas and corresponding labels (urban or rural) it can build the capacity to classify new unlabelled aerial images as either urban or rural. Features in the above example would be different components of the aerial images (i.e. pixel tone and locations) and the target variable would be the label (i.e. urban or rural). When a precise value is required as opposed to a label, it is called a 'regression task'(e.g. Bishop, 2006). An example of this is in flood damage modelling, where features such as water depth, flow velocity and building materials can be used to predict a target variable such as monetary economic damage based on historical records (e.g. Merz et al., 2013; Wagenaar et al., 2017). Due to the use of labelled training data (e.g. classified images or historic damage examples), regression and classification are called supervised learning tasks. Machine learning method categories also include unsupervised learning and reinforcement learning, (see GFDRR, 2018). However, such methods are not discussed in this paper because they are expected to have a smaller short-term impact on the field of flood risk and impact assessments.

The simplest machine learning algorithms have been used for a long time and are often known as basic statistical techniques (e.g. linear regression: Legendre, 1805;

Gauss, 1809). More sophisticated machine learning techniques that emerged in the 1980s and 1990s (e.g. Decision Trees and Neural Networks) can find more complex non-linear patterns (Breimann et al., 1984; Rumelhart et al., 1986). Recent advances in machine learning (e.g. convolutional neural networks) make computer vision and other advanced applications possible (Krizhevsky et al., 2012). The more advanced techniques such as decision trees, neural networks and especially convolutional neural networks can find more complex patterns. This is because they allow for more complex non-linear functions to be fitted to the data. Such complex functions require a large number of model coefficients to be set during the training of the model. To set all these coefficients a lot of training examples are required. In some cases the number of training examples can be reduced with transfer learning techniques (Olivas et al., 2010). These techniques make it possible to re-use knowledge gained from other problems to train a model on a smaller training data set.

From the beginning, machine learning has been used in predictive flood hazard modelling (Solomatine & Ostfield, 2008) mostly as a faster and simpler alternative to process models. A simple example of this is the prediction of river discharge based on upstream rainfall data (e.g. Dibike & Solomatine, 2001). This type of modelling has been practiced for a long time but hasn't displaced the traditional process models. This is probably because the methods aren't sufficiently better than traditional methods to offset some disadvantages as discussed in the predictive hazard section. In recent years, more data is becoming available through remote sensing, social media (e.g. Fohringer et al., 2015), citizen science (e.g. Annis & Nardi, 2019) and other sources. This impulse of new data combined with machine algorithms could lead to changes in flood risk and impact assessment. Some of these changes have already been highlighted by major international organizations such as the World Bank and others (GFDRR, 2018).

This invited perspective paper starts with a perspective per risk assessment component as defined in table 6.1. These specific perspectives start with a description of the traditional approach for the assessments. Followed by a literature review on how machine learning techniques are currently being developed to improve the traditional approach and then proceed to speculate on potential future improvements. This is followed by a general perspectives section in which general trends that come back in the different components are identified and discussed. This includes common challenges (i.e. data limitations, transferability, ethics and bias) and ends with some speculation about the likelihood of future developments.

6.2. Perspective per component

6.2.1 Exposure assessment

6.2.1.1 Descriptive exposure assessments

Descriptive exposure assessments consist of detecting and characterizing (spatial) features such as current buildings and infrastructure, agriculture fields, roads and other infrastructure. Traditionally this has been done by population censuses, building counts and conventional mapping techniques that require ground surveys. Remote sensing is currently changing this. It has become common that aerial and satellite images are being manually digitized and labelled to make building footprints or map roads. This has been done by "crowds" of mappers in "mapathons", for example using the OpenStreetMap platform. Machine learning is very likely going to drastically change this. Research to automatically labelling remote sensing data has already been going on for some time (e.g. Heermann & Khazenie, 1992; Giacinto & Roli, 2001). It is currently already being used to label build-up areas based on nighttime lights (Goldblatt et al., 2018) or satellite images (Goldblatt et al., 2016). Furthermore, algorithms are already being used to automatically label buildings (Sermanet et al., 2013; Alshehhi et al., 2017; GFDRR, 2018) and map roads (Gao et al., 2019) using aerial / satellite imagery. This will reduce the need for manual detection and will probably provide global availability of such building footprints and road information in the near future.

Part of an exposure assessment is the observation of asset features relevant for risk analysis. For example, building materials, building occupancy (e.g. residential or industrial), building height, ground floor elevation, poverty rates in the population, etc. This information is typically not available, but could be very valuable as input for impact models (e.g. Merz et al., 2014; Wagenaar et al., 2017; Schröter et al., 2014) or, for example, to account for poverty in cost-benefit analyses (e.g. Kind et al., 2016). Similarly, ground floor elevation information could radically improve urban pluvial flood damage modelling as damage from small-scale floods is very sensitive to such variables.

Some work has already been carried out on detecting poverty (Watmough et al., 2019) and building heights (Saadi & Bensaibi, 2014) by satellite imagery. Another source of this building feature information could be 360-degree street view images combined with computer vision techniques. Such images are available in, for example, the open source streetview data platform Mapillary (Neuhold et al., 2017). Such techniques are already starting to impact earthquake risk assessments, such as in Guatemala, where 360 degree imagery was fed into Mapillary algorithms in order to automatically detect “soft story” buildings; those most likely to collapse in an earthquake. This was done by having the machine learning algorithm detect features that were indicators of large openings on the ground floor of buildings (large doors, garage doors, shop windows, etc.) (GFDRR, 2018). Computer vision techniques from street level imagery are currently limited to detecting such relatively simple features. However, based on recent advances seen in other computer vision applications (e.g. facial recognition), it is likely that in the future it will be possible to detect more complex building features also. For computer vision models to detect complex information like ground floor elevation or building materials, it would be necessary to provide labelled examples to the algorithms. Such labelled examples are sometimes already available for some areas, e.g. ground floor elevation (Bouwer et al., 2017) or building materials (Schröter et al., 2018).

6.2.1.2 Predictive exposure assessments

Predictive exposure mapping consists of estimates of future exposure. This mostly includes modelling to predict urban growth and other changes in land-use. It is required for evaluating flood mitigation measures (e.g. Wagenaar et al., 2019) because such measures typically need to function for a long time and should therefore still perform as required after predicted land-use changes. Land-use changes affect the impact of a flood because more damage may occur for the same flood hazard and the flood hazard may become greater because of changes in impervious area and therefore rainfall-runoff (Triantakonstantis and Mountrakis, 2013; Mestav Sarica et al., 2019). Predictive exposure assessments for flood risk and impact assessments are currently often not carried out spatially, but rather GDP growth projections are applied to estimate future total exposure values (e.g. van der Most et al., 2014; Wagenaar et al., 2019). This is enough for some studies but if large changes are expected a land-use change or urban growth model is required.

Urban growth has been modelled with simple machine learning models in the past (e.g. logistic regression) (Samardzic-Petrovic, 2017). The use of Cellular Automata (CA) models has become more common recently (Naghibi et al., 2016). These models assign cells as either urban or non-urban based on specific transition rules. Determining the optimum transition rules is a critical issue for CA modelling (Aarthi and Gnanappazham, 2019). This is sometimes difficult because of human bias, heterogeneity and nonlinear relations between driving factors and urban expansion (Naghibi et al., 2016; Xu et al., 2019). To overcome these limitations, machine learning algorithms such as artificial neural networks have been integrated with traditional CA to model urban growth (Aarthi and Gnanappazham, 2019; Naghibi et al., 2016). They then use historical land-use changes (e.g. Song et al., 2015) to learn the transition rules. Complex machine learning models have also been directly applied to urban growth modelling without the CA model structure (Pal and Ghosh, 2017). These improvements, together with more data about past land-use changes and additional computation power, are expected to provide better future land-use maps and make high-resolution future land-use maps globally available.

6.2.2 Hazards assessment

6.2.2.1 Descriptive hazard assessments

Descriptive flood hazard assessment focuses primarily on the response phase, i.e. in estimating current inundation extents and depths to assist both emergency responders and those affected directly. This is traditionally achieved using optical remote sensing data, local sensors or manually collected data from observers on the ground. However, the rise of two major data sources, Synthetic Aperture Radar (SAR) and social media, provides a number of opportunities for machine learning to improve upon current flood detection methods.

During a flood event, affected populations frequently produce 'user-generated content' or 'crowd-sourced' data from social media posts or apps where citizens can report floods (Mazoleni et al., 2017; Assumpção et al., 2018; Annis & Nardi, 2019; UrbanRiskLab, 2019). This is especially the case in urban areas where internet and social media penetration are higher compared to rural areas. This data is often 'tagged' temporally and spatially and can be used by machine learning algorithms for applications such as nowcasting by searching for certain keywords like "flood" (e.g. see Tkachenko et al. 2017, Bischke et al. 2017, Lopez-Fuentes et al. 2017). The method is currently used to map real-time flood extents in several countries (Eilander et al., 2016). Potential future machine learning and computer vision techniques could be extended to estimate water depths and other flood characteristics from posted photos.

Remotely-sensed optical data is often used to identify the extents of flooding, but optical sensors are not functional during periods of cloud-cover or at night. Furthermore, the temporal resolution often prevents the observation of flash floods. SAR data using the microwave wavelengths of the electro-magnetic spectrum can help overcome these problems by providing additional imagery during the night or during cloud cover. Adding night-time and cloud-cover images will provide a higher total temporal resolution. Flood extents are currently determined with statistical methods using thresholds to subsequently identify flood extents e.g. by using Bayesian method on SAR amplitude time-series data (Lin et al. 2019). Advanced machine learning classification methods are being developed to improve this process, but in order to train them it is necessary to have manually labelled images as training data. Collection of this labelled flood extent information is the main challenge for automatic detection moving forward. Manual methods could harness the power of the crowd, as people are connected through the internet or with mapathons. These approaches could have game-changing implications for the training of machine learning algorithms. Already mapathons are often 'trainathons', where mappers are not only manual digitizers, but also labellers and trainers for automated machine learning methods for the future.

6.2.2.2 Predictive hazards assessments

Predictive flood hazard assessments consist of predicting future floods and their characteristics such as extents, inundation depths, durations, flow velocities, waves and water levels in rivers or seas. These assessments are applied for short-term forecasting in the preparation phase (preparing for imminent events) and long-term risk analyses for use in flood risk management (mitigation phase).

In flood-forecasting, traditional methods of predicting hazard variables can involve a chain of hydrologic and hydraulic models that describe the physical processes. Although such models provide system understanding, they often have high computational and data requirements. Therefore, the use of process models may not always be feasible or necessary in the preparation stage of a disaster. At that moment, accurate and timely outputs become more important than system understanding, and the use of 'black-box' machine learning models (e.g. Campolo et al., 2009) is becoming more widespread (Mosavi et al., 2018). The increased speed can create a trade-off with the robustness of forecast models, as changes to the hydraulic system (such as a new structure that could be easily implemented into a hydraulic model) cannot be directly introduced into a trained machine learning model. In addition, machine learning models might not perform well in predicting extremes far outside past observations, since it has not been trained against such extremes.

A review of flood forecasting methods using machine learning by Mosavi et al. (2018) highlights trends such as component and ensemble models (collectively termed 'hybrid models', Corzo & Solomatine, 2014). Hybrid component models assign machine learning a specific task in the modelling process that is either highly complex or not well understood. Examples of this include using machine learning for error correctors (see, for example, studies by Abrahart et al. 2007 and Google Research - Nevo et al. 2019) or flows subject to human influence (Yaseen et al., 2019). Hybrid ensemble methods often use machine learning models to supplement process models, providing robust predictions and uncertainty ranges (Solomatine & Ostfeld, 2007). Such methods benefit from the speed and ability to deal with non-linear multi-variable problems of machine learning modelling and the process understanding available in conventional modelling. The review by Mosavi et al. (2018) does not consider gridded / spatial forecasting techniques, but advanced machine learning techniques are starting to be developed for precipitation pattern nowcasting (Shi et al. 2015) and flood extents prediction (Chang et al. 2018). Another application of machine learning in the preparation phase is in the real-time control of flood defences and systems (e.g. Lobbrecht & Solomatine, 2002; Castelletti et al., 2010). For example, Lobbrecht & Solomatine (2002) used machine learning methods to optimise control decisions in the event of communication network breakdowns during extreme storm events.

Another major application for machine learning in long-term risk analysis is ‘surrogate’ modelling (Ong et al. 2008), in which the outputs from process models are used to train computationally less-intensive machine learning models. This can be applied to speed up different types of process models applied in predictive hazard modelling. For example, in flood defence analysis and design, classical reliability techniques such as First Order Reliability Methods (FORM) and Monte-Carlo simulations (Steenbergen et al. 2004), or large-scale risk analyses that utilise them (Curran et al. 2019), can be replicated using a relatively small amount of evaluations as samples (Chojaczyk et al. 2015, Kingston et al. 2011). However, surrogate models may be particularly susceptible to extrapolation problems, where input data outside the range of the training data is introduced (Ghalkhani et al., 2013).

In the mitigation phase a chain of hydrologic and hydraulic models that describe the physical processes is typically applied (e.g. Wagenaar et al., 2019). In general, system understanding is required to assess proposed or potential future changes. In such cases, data-driven approaches are typically not applicable as there is no data about how the system behaves after the changes occur and hence simulation models are required that describe the physical system.

6.2.3 Flood impact/vulnerability assessment

6.2.3.1 Descriptive impact/vulnerability assessments

Descriptive impact assessments consist of making estimates of the flood impact after or during an event. This is traditionally done with manually collected data from observers on the ground. However, such manual ground inspections are slow and require people to enter the disaster area. Remote sensing can be used to get a very quick first impression of the damage to help with disaster response. Such techniques have already been applied, for earthquake and wind damage (e.g. Menderes et al., 2016). For flooding, this is often more difficult because damage inside buildings is difficult to obtain either from aerial-based or space-based sensors. Only when buildings completely collapse or are removed by strong flows does remote sensing become feasible. This is, for example, the case with flash floods, tsunamis or some storm surges. 360-degree streetview images collected after a flood could potentially be used for damage assessment. Machine learning techniques could then eventually be used to give a quick first estimate of the damage.

The use of machine learning techniques for automatic detection of damages from remote sensing information (aerial or streetview) requires labelled training data from manually collected data from observers on the ground. This data is currently rare. An approach could be to start using remote sensing data to manually label the impact. A way to get around this limitation is to detect changes in pre- and post-flood using high-resolution satellite images for urban areas where many buildings are damaged. Pixels with changed information will denote the damage that happened due to the floods. Eventually this data can then be used as training data for cases where only the post-flood images are available within a short time interval after the flood event. This method would however only be relevant for catastrophic floods because it doesn’t address the fact that most damage remains not observable from top-view. On top of that this approach introduces significant new error: (1) error in the change detection signal, (2) error in relating the change to damage, (3) error in training a new model based on those damage labels. Imagery from different angles (e.g. from streetview or drones) might be more useful for change detection, however this data would also be more difficult to acquire.

6.2.3.2 Predictive impact/vulnerability assessments

Predictive flood impact assessments include models that translate hazard and exposure information into socio-economic impacts of the flood using knowledge on vulnerability. This can include information such as monetary flood damage, casualties, buildings damaged, crop damage, disease outbreak, building materials needed, recovery time, health monitoring of key structures or indirect damage (damages that occur in a different spatial and/or temporal setting than the originating event).

Most predictive flood damage modelling relies on depth-damage functions that describe a relationship between the water depth and monetary flood damage (Merz et al., 2010). They are either based on historical flood damage records (e.g. Thieken et al., 2008; Kreibich et al., 2010) or on expert estimates (e.g. Penning-Rowsell et al., 2005). In practice, many more variables than water depth have an influence on the flood damage (Cammerer et al., 2013, Wagenaar et al., 2016). Therefore, in the scientific literature there has been a move towards multi-variable flood damage models that use many variables (e.g. flood duration, velocity, building materials, socio-economic status of inhabitants etc.) instead of just water depth (e.g. Merz et al., 2013; Spekkers et al., 2014 Chinh, 2015; Kreibich et al., 2017; Wagenaar et al., 2017; Carisi et al., 2018; Amadio et al., 2019). These models are based on data and machine learning. The problem lies with insufficient data availability to train machine learning models and that using the models requires a lot of feature data about flood and building characteristics plus socio-economic data about inhabitants (Wagenaar et al., 2017). In the future we expect more data about features to become available from computer vision applied to street view, satellite or drone images (see descriptive exposure section). This would improve the quality of such models, could make it easier to apply them and make the development possible for more areas.

Machine learning could also be applied to predict disease outbreak after floods by combining remote sensing, meteorological, and socio-economic data (e.g. Mayfield et al., 2018; Carvajal et al., 2018; Modu et al., 2017; Yomwan et al., 2015; Tiwari et al., 2013; Shively et al., 2015). In a flood event, there is an increased risk of infectious diseases among survivors and displaced persons such as measles, diarrhea, acute respiratory infections and malaria can be responsible for many deaths (Lignon, 2006). Predictive modelling of such diseases is rarely carried out, and current approaches mostly focus on simple regression models or process models that simulate the spread of pollutants in the water. One major challenge is that the degree to which such epidemics occur, is associated with the regional endemicity of specific diseases, the nature and scope of the disaster, the level of public health infrastructure in place both before and after the event, and the level and efficacy of disaster response (Ivers & Ryan, 2006). Machine learning models could take such complex processes better into account.

Machine learning can be used for structural health monitoring, this has applications in the preparation phase (Pyayt et al. 2014, Jonkman et al. 2018) and in the long-term reliability analysis required in the mitigation phase (Prendergast et al. 2018, Klerk et al. 2019). Structural health monitoring is in the preparation phase often done by manual inspections of the infrastructure on the ground, for example in the Netherlands there is a large network of volunteers that can be activated in case of high river levels to inspect the dikes. In the mitigation phase this is done by geotechnical process models fed by observations from the ground (e.g. De Waal, 2016). This is for example applied to decide on dike strengthening. Machine learning algorithms can help detect damage patterns from sensor data and are currently being used for the monitoring of flood defence structures such as dikes (Pyayt et al. 2011). Similar methods have also been applied to bridges (Neves et al. 2017). The use of both machine learning algorithms and traditional techniques for damage detection during floods is still very scarce (Prendergast et al., 2018, Pyayt et al., 2011); however, integration of structural health monitoring with flood early warning systems is a very promising field of development for machine learning techniques but would also requiring training data.

Indirect damages and business interruption are often taken into account simply through a scaling factor of the direct damage (e.g. Wagenaar et al., 2019). More complex models to quantify such damages include input-output models and general equilibrium models (e.g. Koks et al., 2016). To quantify indirect damages, such as business interruption losses, estimating the time it will take for different assets to be back in full or partial functionality is required. These post-disaster restoration models have started to be formalized in the last few years, primarily focused on earthquake disasters (Kang et al., 2018; Burton et al., 2018). Due to a lack of gathered empirical data on post-disaster recovery, the use of data-intensive machine learning techniques has not yet made an impact on this discipline. However, the need of probabilistically quantifying recovery will require the use of statistical models for calibration or assessments of recovery times, and that might be possible in the near future with the use of new remote sensing and crowd-sourcing technologies to obtain the empirical feature data needed.

6.3. General Perspectives

6.3.1 Data limitations

Many machine learning applications in flood risk and impact modelling appear to be limited by a lack of data, especially training data needed to build effective machine learning models. This is especially true since the field of flood risk analysis is concerned primarily with extreme events, which are rare, and data-collection during such events is often logistically difficult. The increase in the amount of data around the world does not necessarily imply that this problem will be resolved in the future. Some data is simply not collected or there are measurement definition or quality issues. To fulfil the potential of machine learning, new data collection efforts will be required, along with data standardization protocols. This will take collaboration between different organisations and stakeholders, setting of data standards and a willingness to share. This problem is common to impact data collected (see 6.2.3.1 and 6.2.3.2), labelled flood extent data (see 6.2.2.1), social media hazard data (see 6.2.2.1) and first floor elevation data (see 6.2.1.1).

6.3.2 Transferability of data

A critical assumption behind machine learning techniques is that the data being used to train a model is representative of the situation the model needs to be applied in. For example, a dataset on damage to concrete buildings is not fully applicable to modelling the damage to thatched huts. It is therefore important to collect heterogenous datasets that cover a large spectrum of potential situations (Wagenaar et al., 2018). Data that isn't fully applicable can still have some value, for example through domain adaptation or transfer learning (GFDRR, 2018) but applicable data is always required as well. Wagenaar et al. (2020) showed that sample selection bias correction, a form of domain adaptation, helps to improve machine learning impact models in a transfer setting. Furthermore, it is important to work on efficient ways to communicate the applicability of data-driven models.

6.3.3 Ethics and Bias

Significant attention is currently being given to questions of the ethics and bias of machine-learning systems across a variety of domains, including facial recognition (Keyes, 2018), automated weaponry (Suchman et al., 2016), criminal justice (Eubanks, 2018) and search engines (Noble 2018). A number of technology companies and research institutions have developed guidelines for evaluating machine-learning systems, but this work is still evolving. Despite similar potential for negative impacts of these tools in flood risk management (Soden et al., 2019), the community has not given these issues as much attention. Such concerns include the potential for reinforcing existing social inequalities and the reduced role of human judgement in modelling processes. These are risks that need to be weighed seriously against the potential benefits of machine learning and explored in greater detail

Biases in machine learning can occur because of datasets that, for a number of reasons, do not fully represent the phenomena which they are meant to describe (e.g. people are accidentally excluded). For example, we often measure what we have data for, rather than measure what matters most, or use training datasets that reinforce past problems. For example, if certain settlements aren't detected in exposure maps, because they use different construction practices than the settlements used in training datasets, they may not receive emergency aid in the event of a flood. These problems can be mitigated by ensuring modelers understand the context of what they are attempting to model. Other ethical issues raised by machine learning in the flood management context include data ownership, transparency, consent, and privacy. For example, some people may object to having their home labelled "vulnerable" on a vulnerability map used by first responders. Privacy concerns may be aggravated by machine learning and other big data techniques. Ethics problems should be addressed by carefully weighing the benefits of collecting certain data against the related privacy costs, in collaboration with people who may be affected by the outcomes of decisions based on machine-learning tools.

An additional ethical concern regarding machine learning in flood risk assessment is misuse of models. In some sectors great advances have been made with machine learning (e.g. facial recognition, self-driving cars). This success for some tasks can lead to an awe-inspiring general attitude towards the techniques (Ames, 2018; Narayanan, 2019). This hype sometimes leads to unwarranted trust in the techniques for tasks machine learning is not (yet) suitable for (Narayanan, 2019). For example, many companies are currently using machine learning for hiring decisions despite well-documented failings of these tools. (Narayanan, 2019; Raghavan et al., 2019). In order to avoid such misuse in flood risk assessment, it is important that machine learning implementations are transparent and supervised by independent machine learning and flood risk assessment experts.

Importantly, flood risk assessments are highly data reliant, and the increased attention to questions of ethics and bias in machine learning systems might serve as an opportunity to drive conversations in our field about the limits of disaster data more broadly. Many of the sources of bias or ethical concerns in machine-learning systems originate in, or share common roots with, other kinds of data used to understand disaster risks .This includes issues such as 1) property values driving what areas gets protected, 2) privacy concerns (which may be aggravated by ML and other big data techniques), 3) how the lack of gender/age/ethnicity disaggregated data on disaster risk masks differential vulnerabilities, and 4) the importance of public participation and the voice of residents of areas portrayed by models as "at risk". Detailed analyses of specific cases (e.g. Soden & Kauffman, 2019) are urgently needed to make further progress in understanding the consequences of the assessment methods we use to understand disasters.

6.3.4 Future Outlook

In the following section we draw some general conclusions about how machine learning will change flood risk and impact assessments. Table 6.2 provides an overview of these predictions.

Table 6.2: Future outlook

	PREDICTIVE	DESCRIPTIVE
EXPOSURE	Likely incremental changes, e.g improved Cellular Automata transition rules	Very likely significant changes e.g. automatic exposure detection including building features
HAZARD	Diverse field, changes are more likely to be complementary or to specific components of modelling	Likely changes in detection due to remote sensing and social media algorithms.
IMPACT	Potential for significant changes (i.e. multi-variable data-driven methods)	Significant changes likely for some elements others will probably remain the same

6.3.4.1 Very likely changes

A few of the trends seem inevitable, primarily in cases where recent technological advances or data that recently became available make next steps obvious. A good example of this is the automatic detection of building footprints and roads from high resolution remote sensing imagery (see 2.1.1). This is already possible and will, especially in data-poor areas, drastically improve the quality of the first response and risk calculation. Further advances in the use of machine learning in descriptive hazard assessment through social media are also inevitable (see 2.2.1), given the amount of data available to social media companies.

6.3.4.2 Likely and potential changes

This is the category that can be shaped the most by individual innovators and the majority of the advances discussed in this paper fall under this category. In this case, the innovation still experiences some kind of obstacle that prevents widespread application. It is typically difficult to predict whether such obstacles can be truly removed in the future and how long that will take. Because flood risk and impact assessments are a relatively small field, the obstacles are often economic feasibility that is difficult to assess combined with conservative users. An example of this is the large-scale collection of impact data which is required for both descriptive and predictive impact modelling (see 2.3.1 and 2.3.3) or the training data required for descriptive hazard assessments (see 2.2.1). Sometimes the obstacle is also technical feasibility, for example whether it will really be possible to extract first floor elevation levels from streetview (see 2.1.2). Innovations are also interdependent, for example, when building feature information can be automatically extracted from streetview, impact models will become easier to train and easier to run and it will make more sense to start collecting the required impact data.

6.3.4.3 Unlikely changes

For some processes, machine learning may not be the best solution from a theoretical perspective. For example, the processes of how water flows are very well known and can be well approximated with existing equations. It, therefore, does not always make sense to pick a machine learning approach. Another situation when machine learning is not applicable is when a system is being modelled on which predictions need to be made that cannot have been seen in the data or when we know from an exploratory data analysis that we have no data for it (GFDRR, 2018). For example, how a system may behave under never seen discharges or after new infrastructure has been built (e.g. new dam in the river). In these cases, machine learning may play a role in some components of the model, but process models will very likely remain crucial in simulating the never before seen conditions. Especially for predictive hazard models (see 2.2.2), there are many elements that are unlikely to change with the advance of machine learning.

6.3.4.4 New practices in flood risk and impact assessments

Most changes to flood risk and impact assessments discussed in this manuscript relate to better models. Such cheaper, faster and more accurate models could possibly yield new practices in flood risk and impact assessments. Cheaper models would make flood risk and impact assessments feasible to carry out for a larger group of users and are therefore likely to make emergency aid and investments in mitigation measures more efficient. Faster methods may speed up emergency response and recovery, especially when manually collected data from observers on the ground are replaced by remote earth observation. More accurate models may lead to more early actions being feasible (Coughlan de Perez et al. 2014) and hence early actions can be carried out that couldn't be carried out before. For example, more targeted measures during the preparation and response phase of a flood. Such new measures include providing emergency payouts even before the event to the most vulnerable people (e.g. Reuters, 2019), prioritization of emergency measures in buildings, targeted disease outbreak prevention (Coughlan de Perez et al. 2014), early shipping of the right emergency goods (Coughlan de Perez et al. 2014) and prioritization of early harvesting of crops.

Acknowledgement

This invited perspective paper benefited from the unique intellectual environment created and facilitated through the Understanding Risk Field lab on urban flooding in Chiang Mai, Thailand in June 2019 (https://urfieldlab.com/). We would like to thank the organizers of the event (Robert Soden, David Lallemant, Perrine Hamel, Katherine Barnes, and Giuseppe Molinario) for providing the unique setting to make this paper possible. We would also like to thank the other participants who provided valuable input on this paper by participating in some of the discussions, namely: Gautam Dadhich, Carmen Acosta, Maricar Rabonza, Pamela Cajilig, Rahul Sharma and Wahaj Habib. Furthermore, we would like to thank the editor (Dr. Heidi Kreibich) and the reviewers (Dr. Fernando Nardi and an anonymous reviewer) for their contributions to the paper. This project is supported in part by the National Research Foundation, Prime Minister's Office, Singapore under the NRF-NRFF2018-06 award.

7. SYNTHESIS

7.1 General findings

Flood damage varies widely across different areas and flood events. This difference is caused by differences in vulnerability, coping capacity and exposure and hazard properties. These risk dimensions vary widely among different countries and areas. This thesis demonstrated that a multi-variable model that considers these risk dimension differences among areas is better able to simulate damage outside of its initial context (Chapter 4).

In this thesis, the two most important performance metrics for models are the mean absolute error and mean bias error. Mean absolute errors indicate the model performance at specific objects; whereas, mean bias errors indicate the general over- or underestimations of the total sum of objects. When a model is not transferred (i.e. based on data from the exact same sample as it is applied to), the mean bias error is typically very small (close to zero) because overestimations cancel out underestimations. When a model is transferred, the mean bias error is introduced. The mean bias error metric can be substantially reduced by both the multi-variable approach and the sample selection bias correction methods presented in this thesis (see Chapters 3 and 5). In Chapter 4, we demonstrated that a multi-variable model helped to limit the mean bias error increase introduced due to a model transfer.

However, both the traditional single-variable damage functions and the multi-variable data-driven methods presented in this thesis seem unable to capture all the complex non-linear damage processes that occur during a flood. This lack is apparent via the mean absolute error metric of these models, which can only be reduced by about 15% (see Chapter 3). Thus, about 85% of the error remains in the model. Considerable further reduction of this absolute error with the current data is likely to be impossible, given that many ML techniques have been attempted and none have had more than a minor influence on this error metric. The only alternative remaining, to reduce this error further using a data-driven approach is to collect more and better data. This option is, therefore, proposed as the best way forward to improve data-driven damage models. The techniques explored in this thesis could have more benefits when they are applied to better datasets.

In the following sections, the main research questions from Chapter 1 are answered:

- What are the main sources of uncertainty in flood-damage models?
- What methods are available to make complex multi-variable damage models?
- Do multi-variable damage models perform well when they are transferred to other locations and events?
- Are there techniques to improve the transferability of multi-variable damage models?
- How can the required data to run data-driven impact models be acquired at scale?

7.2 Specific findings of this thesis

7.2.1 What are the main sources of uncertainty in flood-damage models?

The large number of damage processes in flooding and the types of vulnerable assets means that it is theoretically possible to add more detail to damage models to improve their performance. In practice, however, damage modeling always requires some simplifications of reality, and the challenge is to find a balance between simple models with many uncertainties and more complex models with higher development costs and data requirements. Therefore, it is crucial to have an excellent understanding of the sources of uncertainty in impact models because not all uncertainties are of equal importance for every problem.

For traditional flood-damage models, this analysis of sources of uncertainty is described in Chapter 2. Flood-damage models differ significantly from each other due to the different assumptions behind the depth-damage functions and processes considered (e.g. flow velocity, flood experience, building materials, or flood duration) (Jongman et al., 2012). In Chapter 2, it is argued that these differences in assumptions make it difficult to apply a damage model to another area.

Two types of uncertainties are distinguished in damage models: epistemic and aleatory. Epistemic uncertainties arise from a lack of general knowledge; whereas, aleatory uncertainties are randomness/noise in which the mean and distributions are known. In terms of damage models, epistemic uncertainty relates to the applicability of damage models, while aleatory uncertainty relates more to the details of what happens to a specific object. Regarding error metrics, a high mean bias error indicates more epistemic uncertainty, and a high absolute error combined with a low mean bias error indicates more aleatory uncertainty. For cost-benefit analyses in flood-risk management, it is mostly relevant to reduce the epistemic uncertainty (bias error), because only the sum of all the different object-level impact estimations is required. For impact forecasting or insurance, the aleatory uncertainty may also be important depending on the detail level of the analysis requiring reliable damage estimates for individual objects. This situation could only be potentially reduced by much more detailed models (e.g. data-driven models based on more and better data).

7.2.2 What methods are available to make complex multi-variable damage models?

This research demonstrates that ML techniques can be applied to capture the potentially non-linear damage processes compared with flood damage models that use single-variable damage functions. Data about historical damage are required for ML techniques, but these data do not always include all the explanatory variables. These variables can, however, be added by using hydrodynamic simulations of the historical flood events, as illustrated in Chapter 3. These variables include water depth, flow velocity, flood duration, and frequency of the flood at particular locations. Furthermore, building information can be added based on official cadaster data.

In this thesis, several ML techniques were applied to develop models from historical data: linear regression, regression trees, bagging trees, random forests, Bayesian networks, and neural networks (see Chapters 3, 4 and 5). In Chapter 3, these techniques helped to reduce the mean absolute error by about 15%, compared with a single-variable damage model, and the mean bias error was reduced by about 50%. The random forest model seems to perform best throughout the thesis. This model is relatively robust and does not require substantial tuning. Regression trees and bagging trees are simpler versions of random forests that generally also perform well, but slightly worse than random forests. In Chapter 3, regression trees only achieved about half the error reduction of random forests on the mean absolute error metric. Bagging trees performed similarly to random forests on the mean absolute error metric, but on the mean bias error metric, random forests performed best. Bayesian networks did not perform well when applied to flood damage in the Netherlands (Chapter 3) but did better in applications for German flood damages and model transfer (see Chapter 4). Bayesian Networks seem to be sensitive to the discretization detail and the method applied for the network learning, which was different between Chapters 3 and 4. Neural networks are also sensitive to the many algorithm settings but have the theoretical advantage that they can extrapolate (i.e.

make damage estimates outside the spectrum of the training data). Multi-variable linear regression performed worse than the random forests in Chapter 3 but may be acceptable for cases in which more linear damage processes are dominant.

All ML models have potential problems with overfitting. Overfitting occurs when a model fits so exactly to the training data that the model no longer performs well in new cases. Each ML algorithm has its own method of dealing with overfitting. Bayesian networks avoid overfitting by a discretization that is not too fine, regression trees work with a maximum number of splits, random forests work by randomly leaving variables out in some trees, and neural networks punish the model performance when complexity is added. A user of ML techniques needs to know how the algorithm deals with overfitting, so this can be avoided by correctly tuning it. From the experience in this thesis, and for the application of damage models, random forests are most easy to tune and typically perform well, and neural networks are most difficult to tune. One of the reasons for these findings is that the quantity of data available to train models in this thesis may not have been sufficient to make robust neural networks. Therefore, the neural networks could have been less sensitive to parameter tuning if a larger quantity of data had been available.

7.2.3 Do multi-variable damage models perform well when they are transferred to other locations and events?

For most places around the world, no recent historical damage data are available to train multi-variable damage models. Models, therefore, need to be spatially and/or temporally transferred from another area. This transfer is possible under the specific condition that there are data points in the training data that are representative of the area where the model will be applied. Therefore, for a generally applicable model, it is useful to have data from a wide spectrum of damage cases.

This problem of transferring damage models is, in theoretical ML terms, a sample-selection bias problem. The ML models are based on training samples that are not fully representative of the situation they need to be applied in because these samples come from a limited number of events and are, therefore, not evenly distributed across the spectrum of possible damage cases. Hence, there is sample-selection bias toward events for which training data are available.

A German model was successfully transferred to the Netherlands, but a Dutch model could not be successfully transferred to Germany (Chapter 4). The probable reason for this failure is that the German model was based on data that contain situations similar to the Dutch flood event. However, the German data also contained many examples that did not appear in the Dutch data. Therefore, the model trained on Dutch data was not able to predict damage in the German flood events sufficiently. In other words, the German data cover a larger spectrum than the Dutch data and are more heterogenous than the Dutch data. Therefore, it seems possible to transfer a multi-variable damage model when it is applied to an event that lies within the spectrum of the training data. This case illustrates that, for a model to be transferable to many other cases, heterogenous data, such as those available for Germany, are more important than a large quantity of data, such as those available for the Netherlands.

7.2.4 Are there techniques to improve the transferability of multi-variable damage models?

As stated above, one issue with the transferability of damage models is the sample-selection bias problem. This issue means that the samples that are selected (i.e. are available) are not fully representative of the situation they need to be applied in. To reduce this problem, there are standard methods available in the ML literature. These methods are sample-selection bias-correction methods, also called domain adaptation. Chapter 5 discussed how these methods can reduce both the mean bias (about 50% reduction) and absolute errors (about 15% reduction). These error reductions can be large, especially when the event for which the model is being applied is very different from the training data. All sample-selection bias-correction methods work by considering the predictor variables (e.g. water depth) of the event it needs to model. The method helps to change the weights of training data in such a way that the distribution of the predictor variables in the training data match the distribution of the predictor variables of the event it is applied to. For example, if the event damage mainly relates to buildings with less than 1m water depth, but the training data have samples above 1m water depth as well, the sample-selection bias-correction methods give greater weight to the observations below 1m water depth in the training data.

One problem with the sample-selection bias-correction methods is that they can overstress the importance of single observations, which can lead to overfitting. For example, a case in which only a few water depths larger than 3m are present in the training data, but the event for which damages need to be predicted has many water depths larger than 3m. These few extreme water depths then become very important, and the ML algorithm matches those exact points. However, the algorithm may fail to match the general relation between water depth and damages. For example, the algorithm may assume that high damages only occur at 3.31m and

3.38m water depth, but not at 3.34m. We found that this problem can be reduced by fitting a multi-variable correlated distribution to the result after the sample-selection bias correction and by sampling new training data from this multi-variable correlated distribution. Chapter 5 illustrated that this method can substantially reduce the bias errors for single estimates.

7.2.5 How can the required data to run data-driven impact models be acquired at scale?

Flood impacts are caused by complex non-linear processes that depend on many variables. Machine-learning techniques are fitted to model such processes. However, one of the main problems for applying ML techniques is that they require large datasets, and there is a lack of socio-economic and building data to apply to these models.

The results of the literature review and a brainstorming session with experts from around the world, as presented in Chapter 6, suggest that new technologies could help with future data collection. Impact models rely, largely, on exposure characteristics such as building materials or the socio-economic status of the inhabitants, which are rarely available. This availability could be improved by collecting data using new techniques. One example is the use of 360-degree street view images (GFDRR, 2018), which make it possible to automatically extract building characteristics for large numbers of buildings using computer vision techniques (a combination of image processing and advanced ML techniques).

This approach will only be realized, however, when, at the same time, more and better data are collected about historical damage. This process may take time and, therefore, data-driven impact models are likely to rely strongly on spatial and temporal transfers from the current few available data sources.

7.3 Limitations

7.3.1 Scope limited to data-driven approach

The scope of this thesis in Chapters 3, 4, 5, and 6 is limited to data-driven approaches for damage modeling. The alternative to data-driven models are synthetic flood damage models. These synthetic models have not been considered in this thesis despite being a common current approach for flood damage modelling.

The idea behind this scope limitation is that synthetic models have been around for a long time and therefore a newer data-driven approach may have more opportunities for improvement. Other signs that a data-driven approach might have more potential for improvement are that damage processes seem in some cases to be very complex with many different processes contributing to the damage. it might be difficult to prioritize which damage processes are genuinely significant for each combination of flood type and building type. Therefore, experts who produce synthetic damage models may not always fully grasp how these complex processes impact different building types. However, there are probably also floods where one or a few flood damage process are clearly dominant dominant (e.g. building collapse due to flow velocity). In such cases a synthetic approach could be a more logical approach especially if the variation in different types of buildings is also relatively small for the study area.

This thesis didn't include a systematic and complete comparison between the current synthetic models and data-driven models. Such a comparison would be difficult to set up in a fair and systematic way because the conclusions would probably depend strongly on the case study, the quality of the synthetic model and the quality of the data available to produce the data-driven model.

7.3.2 Data limitations

Chapters 3, 4, and 5 revealed that there are still considerable uncertainties in the predictions made using the multi-variable damage models based on ML as applied in this thesis. The mean absolute error reductions are especially limited. The main reason for this weakness seems to be the limited quality of data, rather than the ML algorithms. This conclusion seems plausible, since the error reductions produced by different ML algorithms are typically limited compared with the total error. The current data have four limitations regarding their application in flood-damage models:

- Errors in historical damage records
- Limited number of explanatory variables
- Limited number of training examples
- Sample-selection bias

Gaps and biases in historical damage records can be caused by differences in definition, recording practices, incentives to increase the damage figures, and hazard modeling errors. Damage depends on its definition, and different damage-collection purposes also require different definitions. For example, some damage records may include increased costs due to a demand surge for recovery resources, while others may not include this. Recording practices can also cause differences. Some damage records in this thesis came from phone interviews with victims, and others from expert visits just after the flood. Such differences may lead to disparities between initial damage estimates and total resulting costs for individual households. Neither method is perfect. Damage may be discovered after the expert leaves, or repairs and replacements may prove more expensive than estimated. Phone interviews rely on the memory, the reliability, and the accounting practices of the interviewed victims. Furthermore, Chapter 3 illustrated that historical damage records are often gathered for (public) compensation payouts or other purposes other than damage modeling. These alternative purposes may cause errors in the damage records. For example, when damage is recorded for payouts, the flood victims may have an incentive to report high damage figures, and the inspector or insurance loss adjuster will try to keep the damage assessment low. This dynamic can lead to assertive flood victims being paid more than moderate victims. Some of the datasets (see Chapter 3) rely on explanatory hazard variables gathered from hydrodynamic models. Such models have a limited accuracy as well. In developing countries damage is often collected in Post Disaster Needs Assessments (PDNAs). These are typically carried out by a multitude of humanitarian agencies in different/separate areas. This can also lead to similar harmonization and spatial coverage issues.

The limited number of explanatory variables is a second reason for errors in the models. Many datasets on historical damages are not collected for damage modeling, so they lack the required explanatory variables. Variables such as flooding depth, pollution, and preparation by households may be missing. Other datasets are collected by asking flood victims, who may not be able to accurately answer questions about these variables. Many potentially interesting variables are currently not available, such as building materials, floor types, doorstep height, building condition before the damage, and value of the building contents. These variables, or proxies for these variables, can help to improve flood-damage models. Examples are provided in Chapter 3.

For this study, there is currently **a limited number of training samples** available. More samples would help to find complex non-linear relationships regarding the many different possible damaging processes. The current number of training samples may be too few to capture all the complexity in the damage processes. This issue is especially true for samples with a variety of buildings and flood characteristics. However, in Chapter 4, it was demonstrated that, in a transfer setting, such a variety of samples for several events in different locations and with different characteristics is more useful than having many samples for a single event when the goal is to develop a general applicable model. Its difficult to give exact guidelines on the number of required training samples. However, in general when there are more explanatory variables available and when the range of these variables is larger more training samples are required to cover the entire spectrum. So the greater the spectrum a model is supposed to cover the more training samples are required.

The training samples need to be relevant for the area the model is applied in. Hence, **sample-selection bias** needs to be as small as possible. The sample-selection bias-correction techniques introduced in this thesis (Chapter 5) can reduce this problem but cannot resolve all the problems. For example, if there are no or few samples available in some part of the spectrum, a sample-selection bias-correction technique will not be sufficient to obtain improved results. More samples are required in that specific part of the spectrum to obtain better results. Therefore, additional data collection efforts from events that are similar to the one being modeled are required. Better algorithms and techniques may also be helpful to reduce the sample-selection bias further (see Section 7.5.3, below).

7.4 Recommendations for practical applications

7.4.1 Decision between data-driven and synthetic approach

Each practical damage modelling project needs to decide between a data-driven and synthetic approach. This thesis doesn't provide guidance for such decisions because a systematic comparison between data-driven and synthetic models is outside the scope of this thesis (see 7.3.1). Therefore, no strong recommendations on the choice between data-driven and synthetic models can be made.

However, based on the findings it seems that a multi-variable model transfers better to new areas because its possible to capture multiple damage processes with such models. With each additional variable a multi-variable model becomes more complex and hence it might be relatively expensive to develop multi-variable models with a synthetic approach. When the purpose is to model a large area with many different types of buildings and damage processes it may therefore be cheaper and more feasible to use a data-driven approach when historical damage records are available (e.g. the Philippines case study in chapter 5).

In developed countries, the variation in the building stock is typically smaller than in developing countries. Therefore, the applicability of a damage model might be allowed to be narrower. Furthermore, in developed countries there are also often larger budgets available to develop good local synthetic damage models. Therefore, especially in the absence of good local data, a synthetic approach is probably for some damage modelling projects the best choice. However, for all areas its advised that for future scientific research better data collection is carried out. Such data is both useful for the development of fully data-driven models as for the calibration and validation of synthetic models.

7.4.2 Choice of data-driven model

When one decides on using a data-driven model rather than a synthetic model the next question is what type of data-driven model. This choice depends on the number of variables, the number of training samples and the complexity of the damage processes. A larger more complex dataset benefits from more complex models. In case the dataset is small and less complex it can even be harmful to use an algorithm that is too complex such as a neural network. A random forest is a relatively simple and stable algorithm that did well for the datasets studied in this thesis. The use of more complex multi-variable datasets seems useful in a sample selection bias setting (i.e. a setting in which a model is transferred). However, for the mean absolute error the difference between data-driven models seems small. Unless better data becomes available it is therefore not useful to use more complex data-driven models to reduce the mean absolute error.

7.4.3 Data collection

To address the data limitations, as outlined in Section 7.3.2, new data collection efforts are required (see also 6.3.1). Data collection should be separated into training data and application data. Training data are needed to develop the ML flood-damage models and consist of damage records and explanatory variables. Application data, on the other hand, are required for each building for which damage needs to be estimated, but these data only require explanatory variables.

Especially regarding the training data, data collection probably includes building visits or interviews with victims, because these data require damage values that are difficult to collect in an automated manner (see Chapter 6). These training data need to be heterogenous, so that the range of where the damage model can be transferred to is sufficiently large. This heterogeneity seems more important for transferability than the number of observations and, hence, building visits are more appropriate, since not that many data points are required to make a transferable damage model. For the application data, on the other hand, it is necessary to have data about every building, but it is not required to collect the rarely available historical damage data. Therefore, for application data, it is recommended to use more economically feasible methods of collection, such as government records or automated methods. One promising automated method is 360-degree street view imagery that is automatically analyzed using computer vision (see Chapter 6).

It is crucial to obtain training data samples from different locations. Furthermore, it is important to have a match between the definitions of the explanatory variables in the training data and the application data. Since these future data will need to come from many different sources, it will be necessary to have very strict and clear definitions and measurement guidelines. Moreover, because this data collection will need to occur on a large scale, it will be necessary to set up a coordinating organization. A good step toward such coordination might be the EU working group on loss data from natural hazards (De Groeve et al., 2015).Similar initiatives are being developed by the World Meteorological Organization (WMO) (WMO, 2018) and the United Nations through the Sendai framework using DesInventar (DesInventar, 2020). However, none of these initiatives is currently collecting data on the level of detail that is required for damage modelling.

7.4.4 Modeling other types of impact

The models in this thesis estimate monetary flood damage (Chapters 2, 3, 4, and 5) and the share of buildings that are destroyed given a certain flood event (Chapter 5). Many other types of natural-hazard impact models could also benefit from such techniques. These models include casualty modeling, health impact modeling, modeling recovery/emergency needs, and modeling different types of (business) interruption damage. These techniques could also be applied for different types of natural hazards, such as wind (as illustrated in Chapter 5), earthquakes, heat, droughts, and landslides. All these types of impacts have in common that they depend on many poorly understood complex non-linear processes, which is something that the approaches explored in this thesis have demonstrated being especially good at. Some of these methods can already be used in applied real-world projects (e.g. the typhoon damage model in Chapter 5), while others require additional development in research.

7.5 Recommendations for research

7.5.1 Vulnerability change modeling

For many studies, it is important to not only consider the current flood risk, but also to assess future flood risks. For example, in a cost-benefit analysis for an infrastructure investment, the investment will typically operate for many decades (e.g. Wagenaar et al., 2019). The impact then needs to be assessed for the city or the building stock of the future and not for the current building stock.

Multi-variable damage models seem an interesting solution for assessing future damage as they can include changes in building characteristics. The damage model can be run with the building characteristics of the future. Analyses based on different scenarios regarding how building characteristics will change in the future can be performed using these models. For this approach to work, it will be necessary to include many building characteristics in the model and some observations regarding damage with different (innovative) building styles in the training data. This approach may, therefore, work best in areas that currently contain many older buildings and that are expected to renew their building stock in the near future.

7.5.2 Data-driven modeling for the effect of measures

Multi-variable models can be used also to model the effect of conventional and unconventional measures. Damage models are already being applied to model the effect of risk-reduction measures that lower the water depths (e.g. Wagenaar et al., 2019). However, changes in other hazard variables, such as flood duration, warning time and preparation (e.g. Bubeck et al., 2012), waves, flow velocity, or rise rate, are generally not included in damage models. When a measure changes such an aspect of the hazard, it is currently difficult to estimate the positive influence on the estimated damage. Estimating the impact of such measures is currently done using synthetic models that include the required variable (e.g. USACE, 2012) or specific historical observations (e.g. Bubeck et al., 2012). A multi-variable model that includes such variables can help to calculate the benefits of such measures. The question is whether current data-driven multi-variable models are already sufficiently effective to perform such analyses, or whether it is necessary to work on better data collection first (see Section 7.4.1).

7.5.3 Improve sample-selection bias-correction techniques

The sample-selection bias-correction techniques as applied in this thesis can be improved and further tailored to impact modeling problems. This improvement can be done in two ways: giving weight to variables based on their importance and improving the synthetic data generation. Currently, either all variables are used to determine the weights ascribed in the sample-selection bias correction, or a limited number of variables is used. In practice, all variables are important for the bias correction, but some are more important than others (e.g. water depth and wind speed are the important ones in Chapter 5). A possible improvement would be to include all variables but to give important variables a greater weight when determining the sample-selection bias correction.

Second, the synthetic data-generation technique introduced in this thesis can be improved. This technique was applied in combination with sample-selection bias-correction methods. This combination resulted in a reduction of the prediction errors and helped to reduce overfitting problems related to sample-selection bias-correction techniques applied to small datasets. The method may, however, also degrade the dataset, as it could cause an information loss compared with the original data that were not resampled. This loss happens because the samples are taken from an imperfect statistical model that cannot fully represent all the complexities in the data. The data points sampled from the statistical model may, therefore, no longer contain all the subtle complex relationships between explanatory variables and damage. Alternative methods for the synthetic data-generation technique, as applied in Chapter 5, will, therefore, need to be developed and tested to address this problem. One alternative that could be considered is the use of differential privacy techniques (Khatri, 2017). These techniques add small perturbations to the data to reduce privacy concerns. Recently, Khatri (2017) found that these perturbations work to prevent overfitting also.

7.5.4 Uncertainty contribution of damage models in forecasting setting

Damage models are typically part of a chain of models that are used together, and all components of these chains have their own uncertainties. For permanent risk reduction measures, studies have been carried out that integrate and compare these sources of uncertainty (e.g. De Moel et al., 2012; De Moel et al., 2014). However, in a forecasting setting, where many uncertainties are a function of time, this hasn't been studied. Such information is very important to determine whether a model prediction is accurate enough to make decisions about early actions. Potentially more early actions are possible when uncertainty in damage models is reduced. It would be good if this hypothesis can be tested in a study that looks at the uncertainty contribution in an impact-based forecasting model chain.

REFERENCES

Aarthi, A. D. and Gnanappazham, L., "Comparison of Urban Growth Modeling Using Deep Belief and Neural Network Based Cellular Automata Model - A Case Study of Chennai Metropolitan Area, Tamil Nadu, India," Journal of Geographic Information System, 11, 1-16, 2019.

Abrahart, R. J. and See, L. M.: Neural network modelling of non-linear hydrological relationships, Hydrol. Earth Syst. Sci., 11, 1563–1579, https://doi.org/10.5194/hess-11-1563-2007, 2007.

Ahmad, M., Hippolyte, J., Mourshed, M., Rezgui, Y., 20172. Random forests and artificial neural network for predicting daylight illuminance and energy consumption. Conference: Building Simulation 2017

Ahmad, M.W., Mourshed, M., Rezgui, Y., 20171. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings Volume 147, 15 July 2017, Pages 77-89

Alshehhi, R., Marpu, P.R., Woon, W., Dalla Maru, M., 2017. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing. Volume 130, August 2017, Pages 139-149

Amadio, M., Scorzini, A. R., Carisi, F., Essenfelder, A. H., Domeneghetti, A., Mysiak, J., and Castellarin, A.: Testing empirical and synthetic flood damage models: the case of Italy, Nat. Hazards Earth Syst. Sci., 19, 661-678, https://doi.org/10.5194/nhess-19-661-2019, 2019.

Ames, M. G. (2018). Deconstructing the algorithmic sublime. Big Data & Society. https://doi.org/10.1177/2053951718779194

Annis A. and Nardi F. (2019): Integrating VGI and 2D hydraulic models into a data assimilation framework for real time flood forecasting and mapping, Geo-spatial Information Science.

Apel, H, B Merz, en H. Thieken. „Quantification of uncertainties in flood risk assessments." International Journal of River Basin Management 6:2, 2008: 149-162.

Assumpção, T. H., Popescu, I., Jonoski, A., & Solomatine, D. P. (2018). Citizen observations contributing to flood modelling: Opportunities and challenges. Hydrology and Earth System Sciences, 22(2), 1473-1489.

Axer, T., Bistry, T., Klawa, M., Müller, M., and Süßer, M. (2012), Sturmdokumentation 2011 Deutschland, Deutsche Rück, Düsseldorf, 2012 (in German).

Bachmann, D., Eilander, D., De Leeuw, A., De Bruijn, K., Weerts, A., Beckers, J., 2016. Prototypes of risk-based flood forecasting systems in the Netherlands and Italy. E3S Web of Conferences e3sconf/201 , FLOODrisk 2016 - 3rd European Conference on Flood Risk Management 7. DOI: 10.1051/e3sconf/20160718018

Becker, A.: Maas-modellen 5de generatie: Modelopzet, kalibratie en verificatie. Deltares rapport 1204280-000-ZWS-0011, 2012 (in Dutch).

Berlinet, A., Thomas-Agnan, C., 2004. Reproducing Kernel Hilbert Spaces in Probability and Statistics. Kluwer Academic Publishers. ISBN 978-1-4419-9096-9

BfG (Federal Institute of Hydrology)(2006), Das Hochwasser der Elbe im Frühjahr 2006, BfG-1514, Koblenz, Germany, 2006 (in German).

Billah, M. „Effectiveness of flood proofing domestic buildings." Msc Thesis, UNESCO-IHE and Dura Vermeer, 2007.

Bilmes, J.: Graphical models and automatic speech recognition. The Journal of the Acoustical Society of America 112, 2278 2002; doi: http://dx.doi.org/10.1121/1.4779134

Bischiniotis, K., van den Hurk, B., Coughlan de Perez, E., Veldkamp, T., Guimarães Nobre, G., Aerts, J., 2019. Assessing time, cost and quality trade-offs in forecast-based action for floods. International Journal of Disaster Risk Reduction, Volume 40, November 2019, 101252

Bischiniotis, K., de Moel, H., Van den Homberg, M., Couasnon, A., Aerts, J. C. J. H., Nobre, G.G., Zsoter, E., Van Den Hurk, B., 2020. A framework for comparing permanent and forecast-based flood risk-reduction strategies. Science of The Total Environment, Volume 720, 10 June 2020, 137572

Bischke, B., Helber, P., Folz, J., Borth, D. and Dengel, A., 2017. Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks, arXiv:1709.05932, 2017. https://arxiv.org/abs/1709.05932

Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2

Blöschl, G., Nester, T., Komma, J., Parajka, J., and Perdigão, R. A. P.(2013), The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods, Hydrol. Earth Syst. Sci., 17, 5197–5212, doi:10.5194/hess-17-5197-2013, 2013.

Bouwer, L.M., 2011. Projections of Future Extreme Weather Losses Under Changes in Climate and Exposure. Risk Analysis, volume 33, issue 5. https://doi.org/10.1111/j.1539-6924.2012.01880.x

Bouwer, L.M., Haasnoot, M., Wagenaar, D. and Roscoe, K., (2017). Assessment of alternative flood mitigation strategies for the C-7 Basin in Miami, Florida. Deltares. 1230718.

Breiman, L., Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984). Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. ISBN 978-0-412-04841-8.

Breimann, L., 2001. Random Forests. Machine Learning 45,5-32,2001.

Briene, M., S. Koppert, A. Koopman, en A. Verkennis. Financiele onderbouwing kengetallen hoogwaterschade. NEI, Ministerie van Verkeer en Waterstaat, Rijkswaterstaat DWW, 2002.

Bubeck, P., Botzen, W. J. W., Kreibich, H., and Aerts, J. C. J. H.: Long-term development and effectiveness of private flood mitigation measures: an analysis for the German part of the river Rhine, Nat. Hazards Earth Syst. Sci., 12, 3507–3518, https://doi.org/10.5194/nhess-12-3507-2012, 2012.

Burton, H.V., Miles, S.B., Kang, H., 2018. Integrating Performance-Based Engineering and Urban Simulation to Model Post-Earthquake Housing Recovery. Earthquake Spectra, Volume: 34 issue: 4, page(s): 1763-1785. https://doi.org/10.1193/041017EQS067M

Cabot, C.: Libpgm Python library, pythonhosted.org/libpgm, 2012. website accessed 10-08-2016: http://pythonhosted.org/libpgm/

Cammerer, H., Thieken, A.H., Lammel, J., 2013. Adaptability and transferability of flood loss functions in residential areas. Nat. Hazards Earth Syst. Sci. 13, 3063–3081. doi:10.5194/nhess-13-3063-2013

Campolo, M., Soldati, A., and Andreussi, P.: Artificial neural network approach to flood forecasting in the River Arno, Hydrolog. Sci. J., 48, 381–398, https://doi.org/10.1623/hysj.48.3.381.45286,
2003

Carisi, F., Schröter, K., Domeneghetti, A., Kreibich, H., and Castellarin, A.: Development and assessment of uni- and multivariable flood loss models for Emilia-Romagna (Italy), Nat. Hazards Earth Syst. Sci., 18, 2057–2079, https://doi.org/10.5194/nhess-18-2057-2018, 2018

Carvajal, T.M., Viacrusis, K.M., Hernandez, L.F.T., Ho, H.T., Amalin, D.M. and Watanabe, K., 2018. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. BMC infectious diseases, 18(1), p.183.

Castelletti, A., S. Galelli, M. Restelli, and R. Soncini-Sessa (2010), Tree-based reinforcement learning for optimal water reservoir operation, Water Resour. Res., 46, W09507, doi:10.1029/2009WR008898

Chang, L., Amin, M.Z., Yang, S.N., Chang, F., 2018. Building ANN-Based Regional Multi-Step-Ahead Flood Inundation Forecast Models. Water 2018, 10, 1283; doi:10.3390/w10091283

Charles S. and Wetter, Michael eds.Proceedings of BS2017: 15th Conference of International Building Performance Simulation Association, San Francisco, USA. pp. 1-7

Chatterton, J, E. Penning-Rowsell, en S. Priest. „The Many Uncertainties in Flood Loss Assessments." In Applied Uncertainty Analysis for Flood Risk Management. 2014.

Chawla N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16 (2002) 321–357

Chinh, D., Gain, A., Dung, N., Haase, D., and Kreibich, H.: Multi-Variate Analyses of Flood Loss in Can Tho City, Mekong Delta, Water, 8, 6, https://doi.org/10.3390/w8010006, 2015

Chojaczyk, A., Teixeira, A. P., Neves, L. C., Cardoso, J. B., and Guedes Soares C., 2015. "Review and application of Artificial Neural Networks models in reliability analysis of steel structures," Struct. Saf., vol. 52, pp. 78–89, Jan. 2015.

Cortes C., Mohri, M., Riley, M., Rostamizah, A. (2008). Sample Selection Bias Correction Theory. In ALT, 2008.

Corzo, P.G.A., and Solomatine, D., 2014. Comparative analysis of conceptual models with error correction, artificial neural networks and committee models. EGU General Assembly 2014

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., and Zsoter, E.: Action-based flood forecasting for triggering humanitarian action, Hydrol. Earth Syst. Sci., 20, 3549-3560, https://doi.org/10.5194/hess-20-3549-2016, 2016.

Curran, A., de Bruijn, K.M., Klerk, W.J., Kok, M. Large Scale Flood Hazard Analysis by Including Defence Failures on the Dutch River System. Water 2019, 11, 1732.

Dahm, R., Verhoeven, G., Gehrels, H., Diermanse, F., Wagenaar, D., Dettrick, D., Meijer, D., Vaidya, S., Champathangham, S. and Smithies, C., (2017). Strengthening Disaster Risk Management and Urban Flood Resilience in Lao PDR – Report 2 assessing flood risk and evaluating design options. Deltares. 1230978-002. World Bank Group.

De Bruijn, K. M., Klijn, F., van de Pas, B., and Slager, C. T. J.: Flood fatality hazard and flood damage hazard: combining multiple hazard characteristics into meaningful maps for spatial planning, Nat. Hazards Earth Syst. Sci., 15, 1297–1309, https://doi.org/10.5194/nhess-15-1297-2015, 2015.

De Bruijn, K.M. Resilience and flood risk management: A system approach applied to lowland rivers. PhD Thesis - Delft University of Technology, 2005.

De Bruijn, K.M., Wagenaar, D.J., Slager, K., De Bel, M., & Burzel, A. (2014). Updated and improved method for flood damage assessment. Deltares.

De Groeve, T., Ehlrich, D., Corbane, C., 2015. Guidance for recording and sharing disaster damage and loss data. Publications Office of the European Union. DOI: 10.2788/186107

De Moel, H., and J.C.J.H. Aerts. „Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates." Natural Hazards 58, 2011: 407-425.

De Moel, H., Bouwer, L. M., & Aerts, J. C. J. H. (2014). Uncertainty and sensitivity of flood risk calculations for a dike ring in the south of the Netherlands. Science of the Total Environment, 473-474(1), 224-234. DOI: 10.1016/j.scitotenv.2013.12.015

De Moel, H., N.E.M. Asselman, and J.C.J.H. Aerts. „Uncertainty and sensitivity analysis of coastal flood damage in the west of the Netherlands." Natural Hazards and Earth Systems Sciences 12 (2012): 1045-1058.

De Ruiter, M.C., Couasnon, A., Van Den Homberg, M.J.C., Daniell, J.E., Gill, J.C., Ward, P.J., 2020. Why we can no longer ignore consecutive disasters. Earth's Future Volume 8, Issue 3, March 2020 . https://doi.org/10.1029/2019EF001425

De Waal, J.P., 2016. Basisrapport WBI 2017. Deltares 1230086-002

DesInventar, 2020. Migrating DesInventar to Sendai Framework format. https://www.desinventar.net/migrate_Sendai.html (accessed 15-6-2020)

Deutsches Komitee Katastrophenvorsorge e.V. (DKKV) (Ed.)(2015), Das Hochwasser 2013 - Bewährungsprobe für das Hochwasserrisikomanagement in Deutschland, (Schriftenreihe des DKKV ; 53), Bonn : DKKV, 2015

Dias, P., Arambepola, N.M.S.I., Weerasinghe, K., Weerasinghe, K.D.N., Wagenaar, D.J., Bouwer, L., Gehrels, H., 2018. Development of damage functions for flood risk assessment in the city of Colombo (Sri Lanka), Procedia Engineering 212 (2018) 332–339. https://dx.doi.org/10.1016/j.proeng.2018.01.043

Dibike, Y.B., Solomatine, D.P., 2001. River flow forecasting using artificial neural networks. Physics and Chemistry of the Earth Part B: Hydrology, Oceans and Atmosphere. Volume 26, Issue 1, Pages 1-7.

Diermanse, F.L.M., Carroll, D.G., Beckers, J.V.L., Ayre, R., Schuurmans, J.M., 2012. A Monte Carlo Framework for the Brisbane River Catchment Flood Study. Hydrology and Water Resources Symposium 2014

Diermanse, F.L.M.: HR2006 – herberekening werklijn Maas. Delft Hydraulics Q3623.00, 2004 (in Dutch).

Dietz, H. (1999), Wohngeba¨udeversicherung Kommentar, 2nd ed., VVW Verlag Versicherungswirtschaft GmbH, Karlsruhe, Germany.

DSWD (Department of Social Welfare Development), 2019. Amendment to memorandum circular no.19 series of 2018 on the guidelines in the implementation of the emergency shelter assistance (ESA) for the typhoon "Ompong" – affected households with damaged houses. DSWD Central Office, Philippines.

Duiser, J.A. Een verkennend onderzoek naar methoden ter bepaling van inundatieschade bij dijkdoorbraak. 82-0644 TNO, 1982.

EC: DIRECTIVE 2007/60/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2007 on the assessment and management of flood risks. Official Journal of the European Union. L 288/27, 2007

Efron, B. and Tibshirani, R.J. , 1993. An Introduction to the Bootstrap. Chapman and Hall, New York. https://doi.org/10.1007/978-1-4899-4541-9

Egorova, R., van Noortwijk, J. and Holterman, S.: Uncertainty in flood damage estimation. International Journal River Basin Management 6, nr. 2, 139-148, 2008.

Eilander, D., Trambauer, P., Wagemaker, J., Van Loenen, A., Harvesting social media for generation of near real-time flood maps. 12th International Conference on Hydroinformatics, HIC 2016

Elmer, F., Thieken, A. H., Pech, I. and Kreibich, H.(2010), Influence of flood frequency on residential building losses, Nat. Hazards Earth Syst. Sci., 10, 2145-2159, 2010. doi:10.5194/nhess-10-2145-2010.

Engel, H. (2004). The flood event 2002 in the Elbe River basin: causes of the flood, its course, statistical assessment and flood damages. Houille Blanche 6, 33–36.

Eubanks, V., 2018. Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Falter, D., Nguyen, D., Vorogushyn, S., Schröter, K., Hundecha, Y., Kreibich, H., Apel, H., Theisselmann, F., Merz, B. (2016) Continuous, large-scale simulation model for flood risk assessments: proof-of-concept. Journal of Flood Risk Management, 9(1), 3-21.

Falter, D., Schröter, K., Nguyen, D., Vorogushyn, S., Kreibich, H., Hundecha, Y., Apel, H., Merz, B. (2015) Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain. Journal of Hydrology, 524, 182-193.

Fang, H., Fang, K. and Kotz, S. (2002), The Meta-elliptical Distributions with Given Marginals, Journal of Multivariate Analysis 82, 1–16

FEMA. „HAZUS-MH MR4 Technical Manual." 2009.

Fohringer, J., Dransch, D., Kreibich, H., & Schröter, K. (2015). Social media as an information source for rapid flood inundation mapping. Natural Hazards and Earth System Sciences (NHESS), 15, 2725-2738.

Friedman, N. and Goldszmidt, M.: Discretizing continuous attributes while learning bayesian networks. In Proc. ICML, page 157–165, 1996.

Fuchs, S. , Heiss, K. and Hübl, J.: Towards an empirical vulnerability function for use in debris flow risk assessment. Nat. Hazards Earth Syst. Sci., 7, 495–506, 2007

Ganguly, K.K., Nahar, N., Hossain, B.M.M.,, 2019. A machine learning-based prediction and analysis of flood affected households: A case study of floods in Bangladesh. International Journal of Disaster Risk Reduction, Volume 34, March 2019, Pages 283-294. https://doi.org/10.1016/j.ijdrr.2018.12.002

Gao, X., Klaiber, C., Patel, D., Underwood, J., 2019. AI is supercharging the creation of maps around the world. Tech@Facebook. https://tech.fb.com/ai-is-supercharging-the-creation-of-maps-around-the-world/. Accessed: August 21, 2019

Gauderius, J. and Kind. J.M. (2012). MKBA WV21. Bijlage D: Bewerking schadegegevens. Deltares 1204144-006

Gauderius, J. Verbetering schadebepaling buitendijkse bedrijven Rijnmond-Drechtsteden. Discussiememo waarderingsprincipes schade bedrijven. RebelGroup, 2012.

Gaughan A.E., Stevens FR, Linard C, Jia P and Tatem AJ, 2013, High resolution population distribution maps for Southeast Asia in 2010 and 2015, PLoS ONE, 8(2): e55882.

Gauss, C.F.,1809.. Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum

Gerl, T., Kreibich., H, Franco, G., Marechal, D. and Schröter, K.: A Review of Flood Loss Models as Basis for Harmonization and Benchmarking. PLoS ONE 11(7): e0159791, 2016. https://doi.org/10.1371/journal.pone.0159791

GFDRR. 2018. Machine Learning for Disaster Risk Management. Washington, DC: GFDRR. License: Creative Commons Attribution CC BY 3.0

Ghalkhani, H., Golian, S., Saghafian, B., Farokhnia, A., Shamseldin, A., 2013. Application of surrogate artificial intelligent models for real-time flood routing. Water and Environment journal. Volume 27, Issue 4. doi:10.1111/j.1747-6593.2012.00344.x

Giacinto, G.and Roli, F.: Design of effective neural network ensembles for image classification purposes, Image Vis. Comput., 19, 699–707, https://doi.org/10.1016/S0262-8856(01)00045-2, 2001

Goldblatt, R., Stuhlmacher, M.F., Tellman, B., Clinton, N., Hanson, G., Georgescu, M., Wang, C., Serrano-Candela, F., Khandelwal, A.K., Cheng, W., Balling, R., 2018. Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover. Remote sensing of environment 205(2018) 253-275.

Goldblatt, R., You, W., Hanson, G., Khandelwal, A., 2016. Detecting the boundaries of urban areas in india: A dataset for pixel-based image classification in google earth engine. Remote Sensing, Volume 8, Issue 8

Grossi P., Kunreuther, H., 2005. Catastrophe Modeling: A New Approach to Managing Risk. Springer Science & Business Media, 2005. ISBN 978-0-387-23129-7

Grunthal, G., Thieken, A., Schwarz, J., Radtke, K., Smolka, A., and ¨ Merz, B.: Comparative risk assessments for the city of Cologne – storms, floods, earthquakes, Nat. Hazards, 38(1–2), 21–44, doi:10.1007/s11069-005-8598-0, 2006

Guimarães Nobre, G, Johannes E Hunink, Bettina Baruth, Jeroen Aerts1 and Philip J Ward. (2020) What will the weather do? forecasting flood losses based on oscillation indices. Earth Future, accepted

Hassoun, 1995. Fundamentals of Artificial Neural Networks. MIT Press Cambridge, MA, USA 1995. ISBN:026208239X

Heckerman, D., Geiger, D. and Chickering, D.M., (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. Machine Learning (1995) 20: 197. https://doi.org/10.1023/A:1022623210503

Heckman, J. (1979). Sample selection bias as a specification error. Econometrica, 47, 153–161.

Heermann, P. D. and Khazenie, N.: Classification of multispectral remote sensing data using a back-propagation neural network, IEEE T. Geosci. Remote., 30, 81–88, 1992.

Helton, J.C., en W.L. Oberkampf. „Alternative representations of epistemic uncertainty.” Reliability Engineering & System Safety 85, 2004: 1-10.

Holland, G. J. (1980), "An analytic model of the wind and pressure profiles in hurricanes", Monthly Weather Review, 108(8), 1212-1218

HOWAS 21, website accessed 03-08-2016: http://dx.doi.org/10.1594/GFZ.SDDB.HOWAS21

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Xie, P., 2015. NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) Version 4.5.

IKSE (International Commission for the Protection of the Elbe) (2004). Dokumentation des Hochwassers vom August 2002 im Einzugsgebiet der Elbe. IKSE Report, Magdeburg, Germany.

Islam, K. M. Nabiul, 1997. The impacts of flooding and methods of assessment in urban areas of Bangladesh. Available from Middlesex University's Research Repository.

Ivers, L.C. and Ryan, E.T., 2006. Infectious diseases of severe weather-related and flood-related natural disasters. Current opinion in infectious diseases, 19(5), pp.408-414.

Jongejan, R., Maaskant, B., 2013. The use of quantitative risk analysis for prioritizing flood risk management actions in the Netherlands. CDA 2013 Annual Conference.

Jongman, B., Kreibich, H., Apel, H. B., Bates, P., Feyen, L., Gericke, A., Neal, J., Aerts, J.C.J.H., Ward, P. Comparative flood damage model assessment: towards a European approach. Natural Hazards and Earth Sciences, 12, 3733-3752. 2012

Jonkman, S.N, 2005. Global Perspectives on Loss of Human Life Caused by Floods. Natural Hazards, Volume 34, Issue 2, pp 151-175.

Jonkman, S. N., Voortman, H. G. , Klerk, W. J., van Vuren, S., "Developments in the management of flood defences and hydraulic infrastructure in the Netherlands," Struct. Infrastruct. Eng., vol. 14, no. 7, pp. 895–910, Jul. 2018.

Jonkman, S.N., Bočkarjova, M., Kok, M., Bernadini, P., 2008. Integrated hydrodynamic and economic modelling of flood damage in the Netherlands. ECOLOGICAL ECONOMICS 66 (2008) 77 – 90. doi:10.1016/j.ecolecon.2007.12.022

Kadaster website. Accessed 25-10-2016: https://www.kadaster.nl/bag

Kang, H., Burton, H., Miao, H., 2018. Replicating the Recovery following the 2014 South Napa Earthquake using Stochastic Process Models. Earthquake Spectra, vol. 34, 3: pp. 1247-1266. https://doi.org/10.1193/012917EQS020M

Kanungo, T., Mount, D.M., Piatko, C., Silverman, R., Wu, A.Y., 2002. An Efficient k-Means Clustering Algorithm: Analysis and Implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 24 , Issue: 7 , Jul 2002)

Kelman I., Spence, R., 2004. An overview of flood actions on buildings. Engineering Geology 73 (2004) 297-309

Keyes, O., 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), p.88

Khan, A., Khan, H., & Vasilescu, L. (2008). Disaster Management CYCLE – a theoretical approach, Management and Marketing Journal, 6, issue 1, p. 43-50, https://EconPapers.repec.org/RePEc:aio:manmar:v:6:y:2008:i:1:p:43-50.

Khatri, A.A.H., 2017. Preventing Overfitting in Deep Learning Using Differential Privacy. State University of New York at Buffalo. PhD Thesis

Kienzler, S., Pech, I., Kreibich, H., Müller, M., Thieken, A. H. (2015): After the extreme flood in 2002: changes in preparedness, response and recovery of flood-affected residents in Germany between 2005 and 2011. - Natural Hazards and Earth System Sciences (NHESS), 15, pp. 505-526. DOI: http://doi.org/10.5194/nhess-15-505-2015

Kind, J. Maatschappelijke kosten-batenanalyse Waterveiligheid 21e eeuw. Deltares - 1204144, 2011.

Kind, J., Botzen, W.J., Aerts, C.J.H., (2016) Accounting for risk aversion, income distribution and social welfare in cost-benefit analysis for flood risk management. WIREs Clim Change2016. doi: 10.1002/wcc.446

Kind, J., R.Vos, A. Tijssen, A. Jeuken and N. Slootjes. Towards the Evaluation of Adaptive Flood Risk Management Strategies for the Rhine estuary – Drechtsteden. 6th International Conference on Flood Management (ICFM6), September 2014, Sao Paulo, Brazil.

Kind, J.M.(2013), Economically efficient flood protection standards for the Netherlands. Journal of Flood Risk Management, 7(2), 103-117, 2013.

Kingston, G. B., Rajabalinejad, M. Gouldby, B. P., and. Van Gelder, P.H.A.J.M, 2011. “Computational intelligence methods for the efficient reliability analysis of complex flood defence structures,” Struct. Saf., vol. 33, no. 1, pp. 64–73, Jan. 2011.

Klemas, V., (2015) Remote Sensing of Floods and Flood-Prone Areas: An Overview. Journal of Coastal Research: Volume 31, Issue 4: pp. 1005 – 1013.

Klerk, W., Schweckendiek, T., Den Heijer, F., Kok, M., 2019. Value of information of Structural Health Monitoring in Asset Management of Flood Defences. Infrastructures 2019, 4(3), 56; https://doi.org/10.3390/infrastructures4030056

Koivumäki, L., Alho, P., Lotsari, E., Käyhkö, A., Saari, A., & Hyyppä, H. Uncertainties in flood risk mapping: a case study on estimating building damages for a river flood in Finland. Journal of Flood Risk Management, 3(2), 166–183. 2010

Kok, M., H.J. Huizinga, A.C.W.M. Vrouwenvelder, en W.E.W van den Braak. Standaardmethode 2005, Schade en Slachtoffers als gevolg van overstroming. HKV, TNObouw, Rijkswaterstaat DWW, 2005.

Koks, E.E., 2016. Economic modelling for flood risk assessment. PhD Thesis – VU University

Koks, E.E., Carrera, L., Jonkeren, O., Aerts, J.C.J.H., Husby, T.G., Thissen, M., Standardi, G., Mysiak, J., 2016. Regional disaster impact analysis: comparing input-output and computable general equilibrium models. Nat. Hazards Earth Syst. Sci., 16, 1911–1924, 2016. doi:10.5194/nhess-16-1911-2016

Koller, D. and Friedman, N.(2009), Probabilistic Graphical Models: Principles and Techniques, MIT Press, ISBN: 978-0-262-01319-2, 2009.

Korndörfer, C., Rümpel, A., Seifert, J. O., Kroll, H., Grebedünkel, D., Jakob, T., Frenzel, F., and Ullrich, H. (2006), Bericht zum Frühjahrshochwasser 2006, Report of the Umweltamt Dresden (Environmental Agency of Dresden), Dresden, Germany, 2006 (in German).

Kreibich H., Botto, A., Merz, B., Schröter, K., (2017), Probabilistic, Multivariable Flood Loss Modeling on the Mesoscale with BT-FLEMO. Risk Analysis Volume 37, Issue 4, 774–787

Kreibich, H., Müller, M., Schröter, K., Thieken, A. H. (2017): New insights into flood warning reception and emergency response by affected parties. - Natural Hazards and Earth System Sciences (NHESS), 17, pp. 2075-2092. DOI: http://doi.org/10.5194/nhess-17-2075-2017

Kreibich, H., Seifert, I., Thieken, A. H., Lindquist, E., Wagner, K., Merz, B. (2011): Recent changes in flood preparedness of private households and businesses in Germany. - Regional Environmental Change, 11, 1, pp. 59-71. DOI: http://doi.org/10.1007/s10113-010-0119-3

Kreibich, H., Thieken, A. H. (2009): Coping with floods in the city of Dresden, Germany. - Natural Hazards, 51, 3, 423-436

Kreibich, H., Seifert, I., Merz, B., Thieken, A. H. (2010). Development of FLEMOcs - A new model for the estimation of flood losses in companies. Hydrological Sciences Journal, 55(8), 1302-1314.

Kreibich, H., van den Bergh, J. C. J. M., Bouwer, L. M., Bubeck, P., Ciavola, P., Green, C., Hallegatte, S., Logar, I., Meyer, V., Schwarze, R., Thieken, A. H. (2014) . Costing natural hazards. Nature Climate Change, 4, 303-306.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS 2012)

Kron, W. (2002) Flood Risk = Hazard × Exposure × Vulnerability. In: Wu, M., et al., Eds., Flood Defence, Science Press, New York, 82-97.

Kron, W. and Ellenrieder, T.(2008), Zunehmende Wetterschäden: Was kostet das die Versicherungswirtschaft? Teil 1: Wetterkatastrophen – Ursachen und Schäden, Korrespondenz Wasserwirtsch., 1, 691–696, 2008 (in German).

Kron, W.(2009), Überschwemmungsüberraschung: Sturzfluten und Überschwemmungen fernab von Gewässern,Wasserwirtschaft, 6, 15–20, 2009 (in German).

Kundzewicz, Z.W., Kanae, S., Seneviratne, S.L., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L.M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, R., Kron, W., Benito, G., Honda, Y., Takahashi, K., Sherstyukov, B., 2014. Flood risk and climate change: global and regional perspectives, Hydrological Sciences Journal, 59:1, 1-28, DOI: 10.1080/02626667.2013.857411

Kunreuther, H., Hogarth, R., Meszaros, J., 1993. Insurer ambiguity and market failure. Journal of Risk and Uncertainty 7, 71-87

Legendre, A.M.,1805. Nouvelles méthodes pour la détermination des orbites des comètes, Firmin Didot, Paris, 1805. “Sur la Méthode des moindres quarrés”

LfU (Bavarian Environment Agency), (2007), Gewässerkundlicher Bericht Hochwasser August 2005, Bayerisches Landesamt für Umwelt, Augsburg, Germany, 2007 (in German).

Liaw, A. and Wiener, M. (2002). Classification and Regression by random Forest. R News 2(3), 18—22

Lignon, 2006. Infectious Diseases that Pose Specific Challenges After Natural Disasters: A Review. Seminars in Pediatric Infectious Diseases. Volume 17, Issue 1. January 2016. Pages 36-45. https://doi.org/10.1053/j.spid.2006.01.002

Lin, Y.N., Yun, S., Bhardwaj, A., Hill, E.M, 2019. Urban Flood Detection with Sentinel-1 Multi-Temporal Synthetic Aperture Radar (SAR) Observations in a Bayesian Framework: A Case Study for Hurricane Matthew. Remote Sens. 2019, 11, 1778; doi:10.3390/rs11151778

Lobbrecht A., Solomatine, D., 2002. Machine Learning in Real-Time Control of Water Systems. Urban Water 4, 2002, 283-289

Lopez-Fuentes, L., Van de Weijer, J., Bolaños, M., Skinnemoen, H., 2017. Multi-modal Deep Learning Approach for Flood Detection. MediaEval’17, 13-15 September 2017, Dublin, Ireland

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A., 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. May 2011.

Martinez-Diaz, L., L. Sidner, and J. McClamrock. 2019. “The Future of Disaster Risk Pooling for Developing Countries: Where Do We Go from Here?” Working Paper. Washington, DC: World Resources Institute. Available online at http://www.wri.org/publication/disaster-risk-pooling.

Mayfield, H.J., Smith, C.S., Lowry, J.H., Watson, C.H., Baker, M.G., Kama, M., Nilles, E.J. and Lau, C.L., 2018. Predictive risk mapping of an environmentally-driven infectious disease using spatial Bayesian networks: A case study of leptospirosis in Fiji. PLoS neglected tropical diseases, 12(10), p.e0006857.

Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M., & Solomatine, D. P. (2017). Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?. Hydrology and Earth System Sciences, 21(2), 839-861.

Menderes, A., Erener, A., Sarp, G., 2015. Automatic Detection of Damaged Buildings after Earthquake Hazard by Using Remote Sensing and Information Technologies. Procedia Earth and Planetary Science. Volume 15, 2015, Pages 257-262. https://doi.org/10.1016/j.proeps.2015.08.063

Merz, B., and A. Thieken. „Flood risk curves and uncertainty bounds." Natural Hazards 51 (Natural Hazards 51), 2009: 437-458.

Merz, B., Elmer, F., Kunz, M., Mühr, B., Schröter, K., Uhlemann-Elmer, S. (2014), The extreme flood in June 2013 in Germany. - Houille Blanche - Revue internationale de l'eau, 1, p. 5-10.

Merz, B., Kreibich, H., and Lall, U., (2013), Multi-variate flood damage assessment: a tree-based data-mining approach, Nat. Hazards Earth Syst. Sci., 13, 53-64, 2013. doi:10.5194/nhess-13-53-2013.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A., (2010), Review article "Assessment of economic flood damage", Nat. Hazards Earth Syst. Sci., 10, 1697-1724, doi:10.5194/nhess-10-1697-2010, 2010.

Merz, B., Kreibich, H., Thieken, A. H., Schmidtke, R. (2004): Estimation uncertainty of direct monetary flood damage to buildings. NHESS, 4, 1, p. 153-163

Messner, F., Penning-Rowsell, E., Green, C., Meyer, V., Tunstall S. and van der Veen A.: Evaluating flood damages: guidance and recommendations on principles and methods. Floodsite report T09-06-01, 2007.

Mestav Sarica, G., Zhu, T. and T.- Pan, C., "Flood Exposure of Shenzhen from Past to Future: A Spatio-Temporal Approach using Urban Growth Modeling," Proceedings of the 7th Annual International Conference on Architecture and Civil Engineering, 400–405, 2019.

Meyer, V., Becker, N., Markantonis, V., Schwarze, R., van den Bergh, J. C. J. M., Bouwer, L. M., Bubeck, P., Ciavola, P., Genovese, E., Green, C., Hallegatte, S., Kreibich, H., Lequeux, Q., Logar, I., Papyrakis, E., Pfurtscheller, C., Poussin, J., Przyluski, V., Thieken, A. H., and Viavattene, C.: Review article: Assessing the costs of natural hazards – state of the art and knowledge gaps, Nat. Hazards Earth Syst. Sci., 13, 1351–1373, https://doi.org/10.5194/nhess-13-1351-2013, 2013.

Modu, B., Polovina, N., Lan, Y., Konur, S., Asyhari, A. and Peng, Y., 2017. Towards a Predictive Analytics-Based Intelligent Malaria Outbreak Warning System. Applied Sciences, 7(8), p.836.

Molinari, D, S. Menoni, G. T. Aronica, F. Ballio, N. Berni, C. Pandolfo, M. Stelluti, G. Minucci (2014) Ex post damage assessment: an Italian experience. Nat. Hazards Earth Syst. Sci., 14, 901-916, doi:10.5194/nhess-14-901-2014

Molinari, D., De Bruijn, K., Castillo-Rodriguez, J., Aronica, G., Bouwer, L.M., 2019. Validation of flood risk models: Current practice and possible improvements. International Journal of Disaster Risk Reduction, Volume 33. https://doi.org/10.1016/j.ijdrr.2018.10.022

Mosavi, A., Ozturk, P., Kwok-wing Chau, 2018. Review: Flood Prediction Using Machine Learning Models: Literature Review. Water 2018, 10, 1536; doi:10.3390/w10111536

Mourad, R., Sinoquet, C., Leray, P.: Probabilistic graphical models for genetic association studies. Brief Bioinform (2012) 13 (1): 20-33, 2011. DOI:https://doi.org/10.1093/bib/bbr015

Naghibi, F., Delavar, M. R., and Pijanowski, B. "Urban Growth Modeling Using Cellular Automata with Multi-Temporal Remote Sensing Images Calibrated by the Artificial Bee Colony Optimization Algorithm," Sensors, 16, 2122, 2016.

Narayanan, A., 2019. How to recognize AI snake oil. https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf (accessed January 27, 2020)

NASA Jet Propulsion Laboratory (JPL), 2013, NASA Shuttle Radar Topography Mission Global 3 arc second. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov)

Nateghi, R., Guikema, S., Quiring, S.M., 2014. Power outage estimation for tropical cyclones: Improved Accuracy with simpler models. Risk Analysis Vol 34, No .6, 2014 DOI: 10.1111/risa.12131

Nateghi, R., Guikema, S.D., Quiring, S.M., 2011. Comparison and validation of statistical methods for predicting power outage durations in the events of hurricanes. Risk Analysis, Vol 31, No. 12, 2011

National Research Council. (2006). Introduction Facing hazards and disasters: Understanding human dimensions. Washington, DC: The National Academies Press. https://doi.org/10.17226/11671.

Nawar, S., Mouazen, A.M., 2017. Comparison between Random Forests, Artificial Neural Networks and Gradient Boosted Machines Methods of On-Line Vis-NIR Spectroscopy Measurements of Soil Total Nitrogen and Total Carbon. Sensors (Basel). 2017 Oct; 17(10): 2428.

Neuhold, G., Ollmann, T., Rota, Bulo, S, Kontschieder, P., 2017. The Mapillary Vistas for Semantic Understanding of Street Scenes. International Conf. on Computer Vision (ICCV) 2017 / July, 2017

Neves, C., González, I., Leander, J., Karoumi, R., 2017 "Structural health monitoring of bridges: a model-free ANN-based approach to damage detection," J. Civ. Struct. Heal. Monit., vol. 7, no. 5, pp. 689–702, Nov. 2017.

Nevo, S., Anisimov, V., El-Yaniv, R., Giencke, P., Gigi, Y., Hassidim, A., Mushe, Z., Schlesinger, M., Shalev, G., Tirumali, A., Wiesel, A., Zlydenko, O., Matias, Y., 2019. Machine Learning for Flood Forecasting at Scale. arXiv:1901.09583

Noble, S.U., 2018. Algorithms of oppression: How search engines reinforce racism. nyu Press.

Olivas, E. S., Guerrero, J. D., Martinez-Sober, M., Magdalena-Benedito, J. R., & Serrano López, A. J. (2010). Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques (2 Volumes) (pp. 1-852). Hershey, PA: IGI Global. doi:10.4018/978-1-60566-766-9

Ong, Y.S., Nair, P., Keane, A.J., 2003. Evolutionary Optimization of Computationally Expensive Problems via Surrogate Modeling. AIAA Journal. Vol. 41. No. 4, April 2003. https://doi.org/10.2514/2.1999

Pal, S. and Ghosh, S. K., “Rule based End-to-End Learning Framework for Urban Growth Prediction,” CoRR, abs/1711.10801, 2017.

Pan, S. J. and Yang, Q., 2010. "A survey on transfer learning", IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345-1359, Oct. 2010.

Papathoma-Köhle, M., Kappes, M., Keiler, M., and Glade, T.: Physical vulnerability assessment for alpine hazards: state of the art and future needs, Nat. Hazards, 58, 645-680, doi:10.1007/s11069-010-9632-4, 2011.

Papathoma-Köhle, M., Zischg, A., Fuchs, S., Glade, T. and Keiler, M.: Loss estimation for landslides in mountain areas – An integrated toolbox for vulnerability assessment and damage documentation. Environmental Modelling & Software 63 (2015) 156-169 , 2015.

Papathoma-Köhle, M.: Vulnerability curves vs. vulnerability indicators: application of an indicator-based methodology for debris-flow hazards. Nat. Hazards Earth Syst. Sci., 16, 1771–1790, 2016.

Pedregosa F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E.: Scikit learn: Machine Learning in Python. Journal of Machine Learning Research 12 2825-2830, 2011.

Penning-Rowsell, E. C., Johnson, C., and Tunstall, S.: The benefits of Flood and Coastal Risk Management: A Manual of Assessment Techniques, Middlesex University Press, London, UK, 2005

Peterman, R.M., en J.L. Anderson. „Decision analysis: A method for taking uncertainties into account in risk-based decision making.” Hum. Ecol. Risk Assess. 5(2), , 1999: 231-244.

Pielke RA Jr, Landsea CW, Musulin RT, Downton M (1999) Evaluation of catastrophe models using a normalized historical record: why it is needed and how to do it. J Insur Regul 18(2):177–194

Pielke, R.A., 2007. Future economic damage from tropical cyclones: sensitivities to societal and climate changes. Phil. Trans. R. Soc. A (2007) 365, 2717–2729. doi: 10.1098/rsta.2007.2086 Published online 30 July 2007

Pistrika A.K. and Jonkman S.N.: Damage to residential buildings due to flooding of New Orleans after hurricane Katrina. Natural Hazards. Vol. 54 Issue 2, pp. 413-434, 2009.

Poussin, J.K., Botzen, W.J.W., Aerts, J.C.J.H. (2014) Factors of influence on flood damage mitigation behaviour by households. ENVIRONMENTAL SCIENCE & POLICY, 40, 69-77, DOI: 10.1016/j.envsci.2014.01.013

Praveena, M., Jaianesh, V., 2017. A literature review on supervised machine learning algorithms and boosting process. International Journal of Computer Applications (0975 – 8887). Volume 169 – No.8, July 2017

Prendergast, L. J., Limongelli, M. P., Ademovic, N., Anžlin, A., Gavin, K., Zanini, M., 2018. “Structural Health Monitoring for Performance Assessment of Bridges under Flooding and Seismic Actions,” Struct. Eng. Int., vol. 28, no. 3, pp. 296–307, Jul. 2018.

Pyayt, A., Mokhov, I., Lang, B., Krzhizhanovskaya, V., Meijer, R., 2011. “Machine learning methods for environmental monitoring and flood Protection,” World Acad. Sci. Eng. Technol., vol. 78, 2011.

QGIS Slope Module, 2020. https://docs.qgis.org/2.8/en/docs/user_manual/processing_algs/gdalogr/gdal_analysis/slope.html. Accessed March 29, 2020

QGIS TRI Module, 2020. https://docs.qgis.org/2.8/en/docs/user_manual/processing_algs/gdalogr/gdal_analysis/triterrainruggednessindex.html . Accessed March 29, 2020

Raghavan, M., Barocas, S., Kleinberg, J., Levy, K., 2019. Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. arXiv:1906.09208. DOI: 10.1145/3351095.3372828

Reichert, P., Omlin, M., 1997. On the usefulness of overparameterized ecological models. Ecological Modelling 95, 289–299. doi:10.1016/S0304-3800(96)00043-9

Reliefweb, 2017. Philippines: Destructive Tropical Cyclones from 2006 to 2016. UN Office for the Coordination of Humanitarian Affair. https://reliefweb.int/map/philippines/philippines-destructive-tropical-cyclones-2006-2016

Reuters, 2019. Bangladesh tries new way to aid flood-hit families: cash up front. https://www.preventionweb.net/news/view/66899

Richert, C., Boisgontier, H., and Grelot, F.: Economic assessment of precautionary measures against floods: insights from a non-contextual approach, Nat. Hazards Earth Syst. Sci., 19, 2525–2539, https://doi.org/10.5194/nhess-19-2525-2019, 2019.

Riggelsen, C.: Learning Bayesian Net-works: A MAP Criterion for Joint Selection ofModel Structure and Parameter. In ICDM, 2008 Eighth IEEE International Conference on Data Mining, pages 522-529, 2008.

Rijkswaterstaat: User's Guide WAQUA: General Information. Version 10.59, October 2013 (in Dutch).

Roos, W. 2003. Damage to buildings. Delft Cluster. TNO Rapport DC1-233-9

Ruig, L., Barnard, P., Botzen, W., Grifman, P., Finzi Hart, J., de Moel, H., Sadrpour, N., Aerts, J., 2019(1). An economic evaluation of adaptation pathways in coastal mega cities:An illustration for Los Angeles. Science of the Total Environment 678, 647-659

Ruig, L., Haer, T., de Moel, H., Botzen, W., Aerts, J., 2019(2). A micro-scale cost-benefit analysis of building-level flood risk adaptation measures in Los Angeles. Water Resources and Economics. https://doi.org/10.1016/j.wre.2019.100147

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning by back-propagating errors. Nature Volume 323. October 1986.

S. Xingjian, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Neural Information Processing Systems, 2015.

Saadi S., Bensaibi, M., 2014. Detection of Buildings height using satellite monoscopic image. 2nd EUROPEAN CONFERENCE ON EARTHQUAKE ENGINEERING. DOI:10.13140/2.1.4985.6005

Samardzic-Petrovic, M., Kovacevic, M. , Bajat, B. and Dragicevic, S., 2017. "Machine Learning Techniques for Modelling Short Term Land-Use Change," ISPRS International Journal of Geo-Information, 6, 387, 2017.

Scawthorn, C., Flores, P., Blais, N., Seligson, H., Tate, E., Chang, S., Mifflin, E., Thomas, W., Murphy, J., Jones, C., and Lawrence, M.: HAZUS-MH flood loss estimation methodology II. Damage and loss assessment, Natural Hazards Review, 7, 72–81, 2006.

Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., Merz, B., 2014. How useful are complex flood damage models? Water Resour. Res. 50, 3378–3395. doi:10.1002/2013WR014396

Schröter, K., Kunz, M., Elmer, F., Mühr, B., Merz, B. (2015): What made the June 2013 flood in Germany an exceptional event? A hydro-meteorological evaluation. - Hydrology and Earth System Sciences, 19, p. 309-327.

Schröter, K., Lüdtke, S., Redweik, R., Meier, J., Bochow, M., Ross, L., Nagel, C., Kreibich, H. (2018): Flood loss estimation using 3D city models and remote sensing data. - Environmental Modelling and Software, 105, pp. 118-131. DOI: http://doi.org/10.1016/j.envsoft.2018.03.032

Scutari, (2010). Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software. July 2010, Volume 35, Issue 3.

Seifert, I., Kreibich, H., Merz, B., Thieken, A. H. (2010) Application and validation of FLEMOcs - a flood loss estimation model for the commercial sector. Hydrological Sciences Journal, 55(8), 1315-1324.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2013. “Overfeat: Integrated recognition, localization and detection using convolutional networks,” CoRR, vol. abs/1312.6229, 2013.

Shively, G., Sununtnasuk, C., Brown, M., 2015. Environmental variability and child growth in Nepal. Health & Place. Volume 35, September 2015, Pages 37-51

Sieg, T., Vogel, K., Merz, B., Kreibich, H., 2017. Tree-based flood damage modeling of companies: Damage processes and model performance. Water Resources Research, Volume 53, Issue 7, July 2017, Pages 6050-6068. https://doi.org/10.1002/2017WR020784

Slager, K, De Bruijn, A. Burzel, K.M., L.M. Bouwer, en D.J. Wagenaar. Verbeteringen gevolgenbepaling van overstromingen in buitendijkse gebieden Rijnmond-Drechtsteden. Deltares 1208261-000, 2013.

Sluijs, L., M. Snuverink, K. van den Berg, en A. Wiertz. Schadecurves industrie ten gevolge van overstromingen. Tebodin. Rijkswaterstaat DWW., 2000.

Snuverink, M., K. Van Den Berg, en L. Sluijs. Schade bij inundatie van buitendijkse industrie. Tebodin, Rijkswaterstaat DWW, 1998.

Soden, R., Kauffman, N. (2019). Infrastructuring the Imaginary: How Sea-level Rise Comes to Matter in The San Francisco Bay Area. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Soden, R., Wagenaar, D. Luo, D., Tijssen, A. 2019. Taking Ethics, Fairness, and Bias Seriously in Machine Learning for Disaster Risk Management. Workshop Paper. Presented at NeurIPS 2019 Workshop on Machine Learning for the Developing World. Vancouver, CA.

Solomatine, D., and Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. Journal of Hydroinformatics (2008) 10 (1): 3–22. doi: 10.2166/hydro.2008.015

Song, X., Sexton, J.O., Huang, C., Channan, S., Townshend, J.R., 2015. Characterizing the magnitude, timing and duration of urban growth from time series of Landsat-based estimates of impervious cover. Remote Sensing of Environment

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., and ten Veldhuis, J. A. E.: Decision-tree analysis of factors influencing rainfall-related building structure and content damage, Nat. Hazards Earth Syst. Sci., 14, 2531-2547, doi:10.5194/nhess-14-2531-2014, 2014

Steenbergen, H.M.G.M., Lassing, B.L., Vrouwenvelder, A.C.W.M., Waarts, P. H., 2004: “Reliability analysis of flood defence systems,”

Suchman, L.A. and Weber, J., 2016. Human-machine autonomies. Autonomous Weapons Systems (pp. 75-102). Cambridge, UK: Cambridge University Press.

Sudderth, E. and Freeman, W.: Signal and Image Processing with Belief Propagation. IEEE Signal Processing Magazine Volume: 25, Issue: 2, 2008. DOI: 10.1109/MSP.2007.914235

Thieken, A. H., Bessel, T., Kienzler, S., Kreibich, H., Müller, M., Pisi, S., Schröter, K. (2016): The flood of June 2013 in Germany: how much do we know about its impacts? NHESS, 16, p. 1519-1540.

Thieken, A. H., Kreibich, H., Müller, M., Lamond, J.(2017), Data collection for a better understanding of what causes flood damage – experiences with telephone surveys. In: Flood damage survey and assessment: new insights from research and practice by Molinari, D., Menoni, S., Ballio, F. (eds.) AGU, Wiley, ISBN 978-1-119-21792-3, Chapter 7, 95-106; DOI: 10.1002/9781119217930.ch7, http://onlinelibrary.wiley.com/book/10.1002/9781119217930, 2017.

Thieken, A. H., Kreibich, H., Müller, M., Merz, B. (2007): Coping with floods: preparedness, response and recovery of flood-affected residents in Germany in 2002. - Hydrological Sciences Journal - Journal des Sciences Hydrologiques, 52, 5, pp. 1016-1037. DOI: http://doi.org/10.1623/hysj.52.5.1016

Thieken, A. H., Müller, M., Kreibich, H., and Merz, B.,(2005), Flood damage and influencing factors: New insights from the August 2002 flood in Germany, Water Resources Research, 41, 2005.

Thieken, A. H., Petrow, T., Kreibich, H., Merz, B. (2006): Insurability and Mitigation of Flood Losses in Private Households in Germany. - Risk analysis, 26, 2, p. 383-395.

Thieken, A.H., Olschewski, A., Kreibich, H., Kobsch, S. and Merz, B.: Development and evaluation of FLEMOps - A new flood loss esimation model for the private sector. WIT Transactions on Ecology and the Environment 118 315 -324, 2008.

Tiwari, S., Jacoby, H., Skoufias, E., 2013. Monsoon babies: rainfall shocks and child nutrition in Nepal. Policy Reasearch Working Paper. The World Bank.

Tkachenko N, Jarvis S, Procter R (2017) Predicting floods with Flickr tags. PLoS ONE 12(2): e0172870. https://doi.org/10.1371/journal.pone.0172870

Tollenaar, D., Dahm, R., Wagenaar, D., (2016). Flood Risk Assessment as input to Informed Decision Making for Future Cities: The case of Bandung, Indonesia. Deltares. 1209884-001. Water Resources Institute.

Triantakonstantis, D., and Mountrakis, G., “Urban Growth Prediction: A Review of Computational Models and Human Perceptions,” Journal of Geographic Information System, 04(06), pp. 555–587, 2013.

Tsamardinos, I., Aliferis, C., Statnikov, A., (2003), Algorithms for Large Scale Markov Blanket Discovery. FLAIRS 2003. American Association for Artificial Intelligence.

Tyralis, H., Papacharalampous, G., Langousis, A., 2019. Review: A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. Water 2019, 11(5), 910; https://doi.org/10.3390/w11050910

UCL (University College London), 2018. Website: http://www.tropicalstormrisk.com

Ulbrich, U., T. Brücher, A. H. Fink, G. C. Leckebusch, A. Krüger, and J. G. Pinto (2003), The central European floods of August 2002: part 1— Rainfall periods and flood development, Weather, 58, 371– 377.

UNISDR, 2018. Economic Losses, Poverty & Disasters 1998-2017. https://www.unisdr.org/2016/iddr/IDDR2018_Economic%20Losses.pdf

UrbanRiskLab, 2019. https://urbanrisklab.org/work#/riskmap/. Accessed: 12-08-2019

USACE, 1997. DEPTH-DAMAGE RELATIONSHIPS FOR STRUCTURES, CONTENTS, AND VEHICLES AND CONTENT-TO-STRUCTURE VALUE RATIOS (CSVR) IN SUPPORT OF THE LOWER ATCHAFALAYA REEVALUATION AND MORGANZA TO THE GULF, LOUISIANA, FEASIBILITY STUDY. Prepared for US Army Corps of Engineers New Orleans District. July 1997.

USACE, 2006. DEPTH-DAMAGE RELATIONSHIPS FOR STRUCTURES, CONTENTS, AND VEHICLES AND CONTENT-TO-STRUCTURE VALUE RATIOS (CSVR) IN SUPPORT OF THE DONALDSONVILLE TO THE GULF, LOUISIANA, FEASIBILITY STUDY. Prepared for US Army Corps of Engineers New Orleans District. March 2006. Contract No. DACW29-00-D-0001

USACE, 2015. Comprehensive Study: Resilient Adaptation to Increasing Risk. Physical Depth Damage Function Summary report. North Atlantic Coast Comprehensive Study (NACCS). January 2015

USACE. Guidelines for risk and uncertainty analysis in water resources planning. Fort Belvoir, VA: Institute for Water Resources 92-R-1, 1992.

Van den Homberg, M., McQuistan, C. (2019). Technology for Climate Justice: A Reporting Framework for Loss and Damage as Part of Key Global

Agreements. In Loss and Damage from Climate Change (pp. 513-545). Chapter 2. ISBN 978-3-319-72026-5

Van den Homberg, M., Meesters, K., Van de Walle, B., 2014. Coordination and Information Management in the Haiyan Response: observations from the field. Procedia Engineering 78 (2014) 49-51

Van den Homberg, M., Monné, R., Spruit, M., 2018. Bridging the information gap of disaster responders by optimizing data selection using cost and quality. Comput. Geosci. 120, 60–72. https://doi.org/10.1016/j.cageo.2018.06.002

Van den Homberg, M., Visser, J., Van der Veen, M., 2017. Unpacking Data Preparedness from a humanitarian prioritization perspective: towards an assessment framework at subnational level, Proceedings of the 14th ISCRAM Conference – Albi, France, May 2017 Tina Comes, Frédérick Bénaben, Chihab Hanachi, Matthieu Lauras, eds.

Van der Most, H., Tanczos, I., De Bruijn, K.M. and Wagenaar, D.J.: New, Risk-Based standards for flood protection in the Netherlands. 6th International Conference on Flood Management (ICFM6), September 2014, Sao Paulo, Brazil. 2014.

Van der Veen, M., 2016, https://www.510.global/philippines-typhoon-haima-priority-index/

Van Lint, S., 2015. "Sense-making of the Netherlands Red Cross Priority Index model – case typhoon Haiyan, Philippines."

Van Ootegem, L., Verhofstadt, E., Van Herck, K. and Creten, T.: Multivariate pluvial flood damage models. Environ. Impact Assess. Rev. 2015, 54, 91–100, 2015.

Van Verseveld, H.: Impact Modelling of Hurricane Sandy on the Rockaways. Relating high-resolution storm characteristics to observed impact with use of Bayesian Belief Networks. MSc thesis Delft University of Technology - Deltares. 2014.

Verkade, J. S. and Werner, M. G. F.: Estimating the benefits of single value and probability forecasting for flood warning, Hydrol. Earth Syst. Sci., 15, 3751–3765, https://doi.org/10.5194/hess-15-3751-2011, 2011.

Vilier, J.J.T.A. 2013. Assessment of the costs of business interruption due to large scale floods: A comparison of methods. Master Thesis Delft University of Technology – HKV.

VNK. Overstromingsrisico dijkring 43 Betuwe, Tieler- en Culemborgerwaarden. Rijkswaterstaat WVL , 2014.

Vogel, K., Riggelsen, C., Kreibich, H., Merz, B. and Scherbaum, F.: Flood Damage and Influencing Factors: A Bayesian Network Perspective, in: Proceedings of the 6th European Workshop on Probabilistic Graphical Models (PGM 2012). Presented at the 6th European Workshop on Probabilistic Graphical Models, Granada, Spain, pp. 347–354. 2014.

Vrouwenvelder, A.C.W.M., Waarts, P.H., 1994. TAW-E rapport Risico Analyse, TNO rapport 94-CON-RO0120, januari 1994

Wagenaar, D., de Jong, J., and Bouwer, L. M.(2017), Multi-variable flood damage modelling with limited data using supervised learning approaches, Nat. Hazards Earth Syst. Sci., 17, 1683-1696, https://doi.org/10.5194/nhess-17-1683-2017, 2017.

Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L.M., Kreibich, H., 2018. Regional and Temporal Transferability of Multivariable Flood Damage Models. Water Resources Research. Volume 54, Issue 5, Pages 3688-3703. https://doi.org/10.1029/2017WR022233

Wagenaar, D.J. The significance of flood duration for flood damage assessment. Master Thesis Delft University of Technology - Deltares, 2012.

Wagenaar, D.J., Dahm, R.J., Diermanse, F.L.M., Dias, W.P.S., Dissanayake, D.M.S.S., Vajja,
H.P, Gehrels, J.C., Bouwer, L.M., 2019. Evaluating adaptation measures for reducing flood risk: A case study in the city of Colombo, Sri Lanka. International Journal of Disaster Risk Reduction. Volume 37, July 2019, 101162.

Wagenaar, D.J., De Bruijn, K.M., Bouwer, L.M. and De Moel, H.,(2016), Uncertainty in flood damage estimates and its potential effect on investment decisions. Natural Hazards and Earth System Sciences, 16(1), 1-14, 2016.

Wagenaar, D.J., Hermawan, T., Van den Homberg, M., Aerts, J.C.J.H., Kreibich, H., De Moel, H., Bouwer, L.M., 2020. Improved transferability of multi-variable damage models through sample selection bias correction. Risk Analysis. Accepted.

Walther, P., Görner, C., Gerber, S., and Wundrak, P. (2013), Ereignisanalyse der Hochwasser von 2010 und 2011 – meteorologische und hydrologische Ursachen, Prozesse, Schäden und Lehren, in: Wasserbaukolloquium 2013: Technischer und organisatorischer Hochwasserschutz – Bauwerke, Anforderungen, Modelle, Dresdner Wasserbauliche Mitteilungen, Dresden, Germany, 48, 141– 148, 2013 (in German).

Watmough, G., R., Marcinko, C.L.J., Sullivan, C., Tschirhart, K., Mutuo, P.K., Palm, C.A., Svenning, J., 2019. Socioecologically informed use of remote sensing data to predict rural household poverty. PNAS January 22, 2019 116 (4) 1213-1218; first published January 7, 2019 https://doi.org/10.1073/pnas.1812969116

White, G. I.: Human adjustments to floods, Department of Geography, University of Chicago, Research Paper No. 29, 1945.

Wind, H.G., Nierop, T.M. ., de Blois, C.J. and de Kok, J.L.(1999), Analysis of flood damages from the 1993 and 1995 Meuse floods. Water Resources Research, 35, 3459-3466. 1999.

WL Delft (1994): Onderzoek watersnood Maas, Deelrapport 1:Wateroverlast December, Deelrapport 9: Schademodellering, Waterloopkundig Laboratorium, Delft, Appendix A (in Dutch).

WL Delft: Onderzoek watersnood Maas, Deelrapport 9: Schade. WL Delft, Delft, 1994. (in Dutch).

WMO, 2015. WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services. WMO-No. 1150. World Meteorological Organization, 2015.

WMO, 2018. Cataloguing High-Impact Events and Associated Loss and Damage. https://unfccc.int/sites/default/files/resource/Excom%207%20submission%20from%20WMO%20(002).pdf (accessed 15-6-2020)

World Bank, 2014. Philippines - Support to Post-Yolanda (Haiyan) Reconstruction and Recovery Planning : Summary of knowledge briefs (English). Report No: ACS11434, October 2014

Xu, T., Gao, J. and Coco, G., “Simulation of urban expansion via integrating artificial neural network with Markov chain – cellular automata,” International Journal of Geographical. Information Science, 33:10, pp. 1960-1983, 2019. (DOI: 10.1080/13658816.2019.1600701)

Yaramakala, S., Margaritis, D., (2005), Speculative Markov blanket discovery for optimal feature. Fifth IEEE International Conference on Data Mining. DOI: 10.1109/ICDM.2005.134

Yomwan, P., Cao, C., Rakwatin, P., Suphamitmongkol, W., Tian, R. and Saokarn, A., 2015. A study of waterborne diseases during flooding using Radarsat-2 imagery and a back propagation neural network algorithm. Geomatics, Natural Hazards and Risk, 6(4), pp.289-307.

Yuzva, K., Botzen, W., Aerts, J.C.J.H., Brouwer, R (2018) A global review of the impact of basis risk on the functioning of and demand for weather index insurance in agriculture. International Journal of Disaster Risk Reduction. 28, 845-853

Zadeh, L.A., 2005. Toward a generalized theory of uncertainty (GTU) - an outline. Information Sciences 172, 1–40.

Zadrozny, B., 2004. Learning and Evaluating Classifiers under Sample Selection Bias. In Proceedings of the Twenty-first International Conference on Machine Learning, ICML ’04, 114–. New York, NY, USA: ACM.

ACKNOWLEDGEMENT

This thesis would not have been possible without the help and support of a great number of people. These people contributed to this thesis, directly through the input they provided and/or indirectly in the support they provided.

I like to thank the Deltares management (Judith Kaspersma and Gerard Blom) for allowing me to spend time on this PhD. Furthermore, I like to thank the Deltares research programs Flood Risk Strategies (Herman van der Most and Alessio Giardino) and Enabling Technologies (Hanneke van der Klis and Jonathan Nuttall) for funding this PhD. I also like to thank the research program Extreme Weather (Ferdinand Diermanse and Robert McCall) for providing extra funding for the fourth paper.

The most important person helping me with this PhD thesis has been my co-promotor Laurens Bouwer. He has been my scientific mentor throughout this process. He contributed heavily on the content. But he also encouraged me, included me in the IMPREX project and provided me with directions and confidence when I needed it. I also really appreciate that he stayed on as a co-promotor after he moved Germany, I also like to thank his new employer Climate Service Center Germany for facilitating that and welcoming me to work at their office. I also like to thank my promotor Jeroen Aerts for giving direction to this PhD thesis, believing in me and for helping me become a better writer. I like to thank Hans de Moel for his curiosity and enthusiasm, I always really enjoy our conversations and his input has been very valuable. Furthermore, I like to thank Bruno Merz, Matthijs Kok, Marc van den Homberg, Albrecht Weerts, Dim Coumou and Gabriella Guimarães Nobre for reviewing this thesis. I also like to thank all the reviewers and journal editors for their input in the papers.

Then, I like to thank all my co-authors on the papers. I like to thank Karin de Bruijn for introducing me to the topic of damage modelling and for trusting and encouraging me from the beginning. I especially like to thank her for improving a lot of my bad writing in the first paper. Jurjen de Jong, I like to thank for a few really good ideas in the second paper. Then I like to thank Heidi Kreibich, Stefan Lüdtke and Kai Schröter for the collaboration in the third paper. I really enjoyed working with them, they are really at the forefront of damage modelling. This thesis wouldn’t have been possible without the foundation coming from them and the rest of GFZ. Then I like to thank Tiaravanni Hermawan, her contributions to the fourth paper were crucial and her friendship was supportive. I also like to thank Marc van den Homberg and Aklilu Teklesadik for our collaboration. The work 510 is doing is very inspiring and I learned a lot from them about how damage models can be applied in a forecasting setting. Then I like to thank the large team that worked on the fifth paper. I like to thank David Lallemant and Robert Soden for organizing the field lab that made this paper possible and for their contributions to this paper, I really enjoy our collaborations. I like to thank Judith Kaspersma and Gerard Blom for allowing me to go to the field lab. The field lab was an amazing experience, I learned a lot there and really enjoyed the brainstorm sessions. Then I like to thank the rest of the team on this paper: Mariano Balbi, Alok Bhardwaj, Emir Hartato, Gizem Mestav Sarica, Laddaporn Ruangpan and Giuseppe Molinario. I really enjoyed our brainstorm sessions and their contributions made this paper possible. But I especially like to thank Alex Curran for this fifth paper, we came up with the initial idea together, we worked on it from the start and I really like to thank him for all his input and for his friendship.

Then, I like to thank the amazing environment in Deltares. Whenever I have a question there is someone in Deltares who has the answer. Related to this PhD, I like to thank Ferdinand Diermanse for answering many math questions, I like to thank Jonathan Nuttall and Hans Korving for their help with the machine learning and Kathryn Roscoe for her help with the Bayesian Networks. However, Deltares also made me a good researcher and gave me a very nice and friendly environment to work in. For that I like to thank many colleagues especially from the Flood Risk Management department.

I like to thank my father, he has been of great importance in the development of my reasoning skills and his grasp of machine learning made me understand the techniques better. I also like to thank my mother, my brothers Robert and Eelco, Linda, Naomi, and my best friend Jordy, their support has been very important to me, especially during the difficult period in 2018. Finally, I like to thank my girlfriend, Nichakarn, her love and support in the last year of this PhD has been crucial. Doing this PhD part-time on a limited budget was very stressful, but every time it got nearly too much, Nichakarn managed to clear my mind, calm me down and got me focused and productive again.

ABOUT THE AUTHOR

Dennis Wagenaar holds a master's degree in civil engineering from Delft University of Technology. He started at Deltares in September 2012. Within the Dutch Deltaprogramme, he worked on the flood protection standards which are now implemented in Dutch law. He also contributed to the standard Dutch flood damage model SSM2017. He was involved in the development of the damage modelling software called Delft-FIAT. Furthermore, he developed flood damage models for Sri Lanka, Afghanistan, Florida, Louisiana, New York, Uzbekistan, Indonesia and Laos. He taught courses on flood impact modelling and disaster risk reduction around the world. He also worked in several EU funded research projects in which he wrote three peer-reviewed scientific journal papers on flood damage modelling. Based on these papers, he started in 2018, a part-time PhD at the VU University in Amsterdam on using machine learning for flood damage modelling. In 2020 he was one of the core organizers of an international working group to produce a guidance note on the responsible use of machine learning in Disaster Risk Management.

Peer-reviewed journal publications

Wagenaar, D.J., Curran, A., Balbi, M., Bhardwaj, A., Soden, R., Hartato, E., Mestav Sarica, G., Ruangpan, L., Molinario, G., and Lallemant, D.,2020. Invited perspectives: How machine learning will change flood risk and impact assessment, Nat. Hazards Earth Syst. Sci., 20, 1149–1161, 2020

Wagenaar, D.J., Hermawan, T., Van den Homberg, M., Aerts, J.C.J.H., Kreibich, H., De Moel, H., Bouwer, L.M., 2020. Improved transferability of multi-variable damage models through sample selection bias correction. Risk Analysis. Accepted.

Sairam, N. Schröter, K., Carisi, F. **Wagenaar, D.J.**, Domeneghetti, A., Molinari, D., Brill, F., Priest, S., Viavattene, C., Merz, B., Kreibich, H., 2020. Bayesian Data-Driven Approach Enhances Synthetic Flood Loss Models. Environmental Modelling and Software. Accepted.

Wagenaar, D.J., Dahm, R.J., Diermanse, F.L.M., Dias, W.P.S., Dissanayake, D.M.S.S., Vajja, H.P, Gehrels, J.C., Bouwer, L.M., 2019. Evaluating adaptation measures for reducing flood risk: A case study in the city of Colombo, Sri Lanka. International Journal of Disaster Risk Reduction.

Dias, P., Arambepola, N.M.S.I., Weerasinghe, K., Weerasinghe, K.D.N., **Wagenaar, D.J.,** Bouwer, L., Gehrels, H., 2018. Development of damage functions for flood risk assessment in the city of Colombo (Sri Lanka), Procedia Engineering.

Wagenaar, D.J., Lüdtke, S., Schröter, K., Bouwer, L.M., Kreibich, H., 2018. Regional and Temporal Transferability of Multivariable Flood Damage Models. Water Resources Research. Volume 54, Issue 5, Pages 3688-3703.

Klijn, F., Asselman, N., **Wagenaar, D.J.**, 2018. Room for Rivers: Risk Reduction by Enhancing the Flood Conveyance Capacity of The Netherlands' Large Rivers. Geosciences 2018, 8, 224

Wagenaar, D.J., de Jong, J., and Bouwer, L. M., 2017. Multi-variable flood damage modelling with limited data using supervised learning approaches, Nat. Hazards Earth Syst. Sci., 17, 1683-1696,

Wagenaar, D.J., De Bruijn, K.M., Bouwer, L.M. and De Moel, H., 2016. Uncertainty in flood damage estimates and its potential effect on investment decisions. Natural Hazards and Earth System Sciences, 16(1), 1-14, 2016.

Other publications

Soden, R., **Wagenaar, D.J.**, Luo, D., Tijssen, A. 2019. Taking Ethics, Fairness, and Bias Seriously in Machine Learning for Disaster Risk Management. Workshop Paper. Presented at NeurIPS 2019 Workshop on Machine Learning for the Developing World. Vancouver, CA.

Van der Most, H., Tanczos, I., De Bruijn, K.M., **Wagenaar, D.J.**, 2014. New, Risk-Based standards for flood protection in the Netherlands. 6th International Conference on Flood Management (ICFM6), September 2014